# Tableau Desktop
# Part 2

# 1 Blending Data to Visualize Relationships

**Lesson Time: 1 hour, 15 minutes**

## Lesson Introduction

In many instances, you can't get all the data you need from a single data source. At those times, you need to combine data from multiple data sources to get the information you need. In this lesson, you will blend data from multiple sources in Tableau® to visualize the data you need and troubleshoot blends so data is accurate.

## Lesson Objectives

In this lesson, you will:

- Blend data from multiple data sources.
- Troubleshoot and refine data blends.

# TOPIC A

## Blend Data

Tableau® offers multiple options to combine data for use in visualizations. In this topic, you will blend data from multiple data sources for use in visualizations.

### Options for Combining Data

***Tableau*** makes it easy to create compelling visualizations from data. One of the most powerful features in Tableau is that it can connect to multiple ***data sources*** simultaneously, allowing you to create visualizations using data from all the connected data sources. There are two ways you can use data from multiple data sources in Tableau:

- Joining data. Joins commonly link tables in the same database, but Tableau provides capability to create cross-database joins to link tables in different data sources. Joins link data at the row level.
- Blending data. Unlike joining, ***data blending*** is used to pull in data from different data sources to get a combined view of information.

Both joins and blends require that relationships are established between tables, worksheets, or data sources using some type of unique information shared among all of them. For example, if pulling product data from multiple worksheets or multiple databases, relationships might be created on the product ID field.

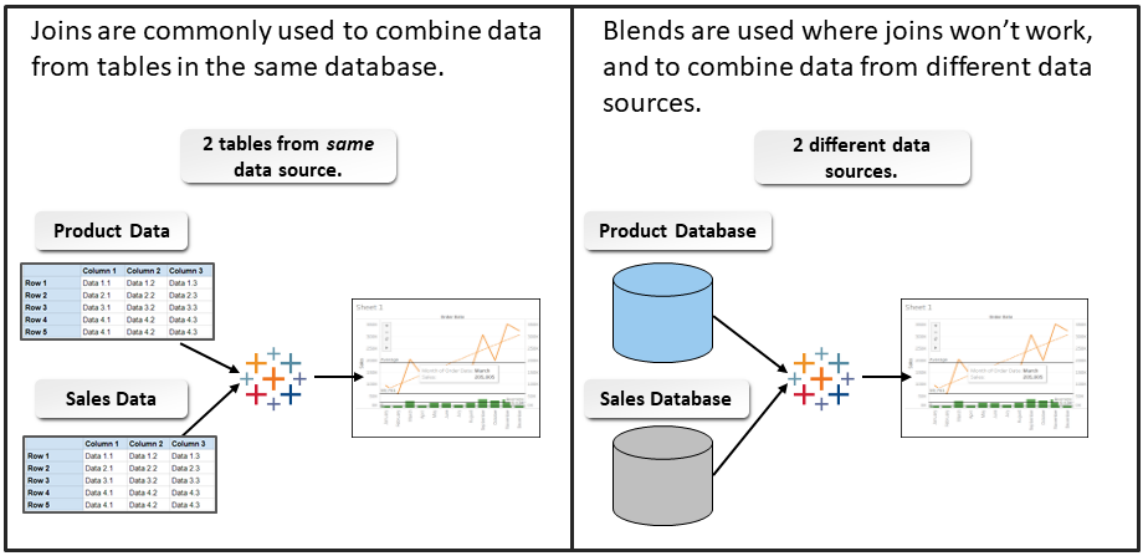



*Figure 1-1: Options for combining data.*

### Data Blending

Data blending allows you to combine data from multiple different data sources into a single view, and include it as part of your data analysis and exploration.

For example, if one data source has data related to products such as size, weight, and cost per unit, and the other has data related to sales such as sales price and shipping costs, then you could use data blending to create visualizations to analyze profitability and the effect of shipping costs on profitability.

To blend data, at least one common field called a ***linking field*** must be present in each of the data sources and a relationship must be established between those common fields. In the aforementioned

example, a blend might be created on the product ID field. Relations may be added automatically or manually and data can be blended on one or more related fields. In the products example, you might blend on both product ID and store identifier if both exist in all data sources to compare product sales and profitability between stores.

When data is blended, data is retrieved by Tableau through queries sent to each data source. Tableau retrieves the query results including data aggregations where it's combined in the visualization.

> **Note:** It's important to note that data blending happens in Tableau after data sources are connected. That means you cannot save a blend as a data source because it's not a data source. Nor can you publish a blend as a single data source. You can save and publish data sources that have been joined.

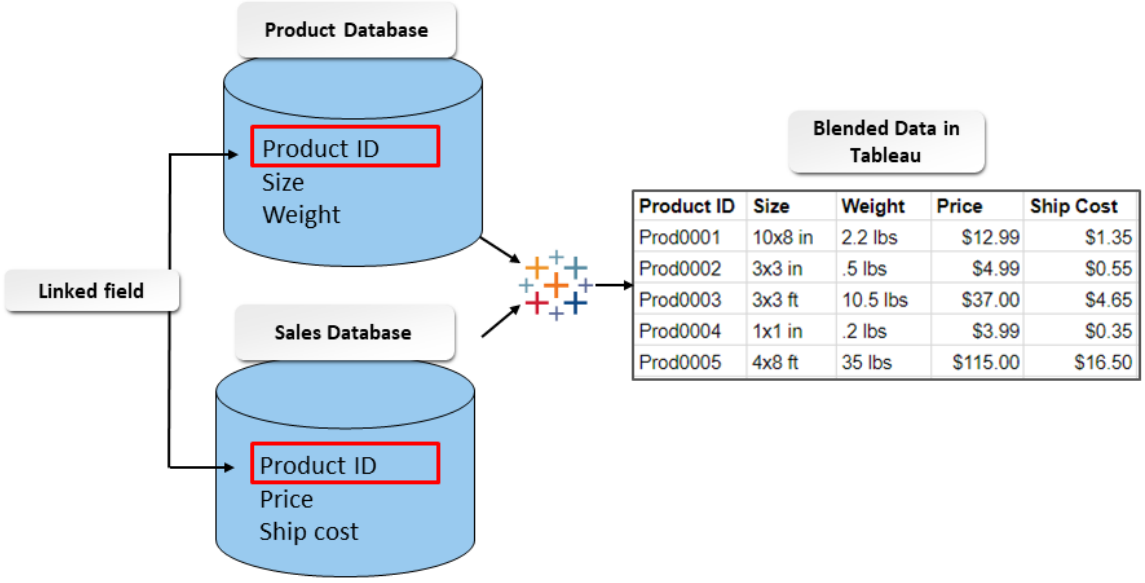



*Figure 1-2: Data blending.*

## When to Blend Data

There are choices that must be weighed when deciding to use data blending versus cross-database joins, but there are some instances where using data blending is mandatory or preferred:

- **When cross-database joins aren't supported**. Cross-database joins aren't supported for multidimensional data sources, such as cubes or for extract-only connections.
- **To simplify the process of combining data that is at different levels of granularity.** For example, if you want to track sales calls made daily versus quotas set monthly for sales staff, both joins and data blending can bring the data into Tableau. However, data blending will aggregate the data at a common level, making it faster to work with the data for analysis and exploration.

## Relationships in Blended Data

The data sources connected are displayed at the top of the **Data** pane. You can select each data source to view the ***dimensions*** and ***measures*** in that data source. To blend data, both data sources must share a linking field between the data sources. If Tableau detects a field with the same name and data type, it will automatically create a relationship establishing a linking field. For example, if two data sources share the field State, a relationship will automatically be established between the State fields, and information blended based on that.

You may need to manually create relationships to get more meaningful results. For example, when looking at product information, if one data source uses the field ProductID, while the other data source uses the field Product_Code, you can manually add a relationship between the two fields

from **Data→Edit Relationships**. You can also rename one of the fields in one data source to match the name of the field in the other data source. To have a valid relationship, fields should contain the same members.

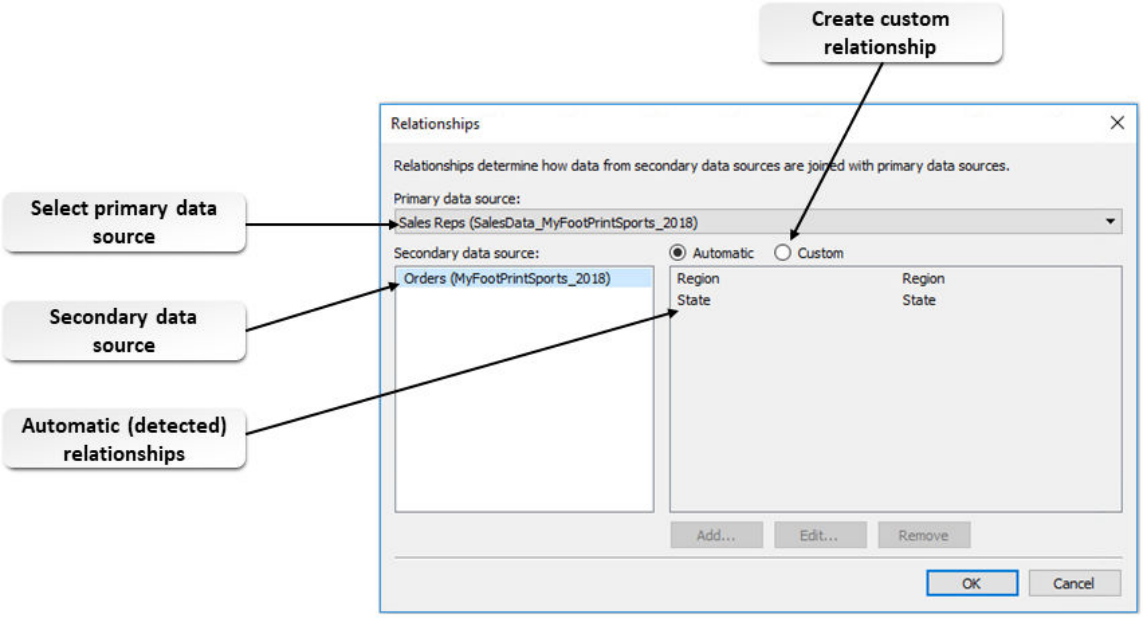



*Figure 1-3: Relationships in blended data.*

## Primary and Secondary Data Sources

In a workbook that has multiple data sources connected through data blending, each ***worksheet*** will have a primary data source and secondary data sources. The data source whose fields are added to the visualization first becomes the primary data source. That is, when the first fields are added to the visualization by double-clicking or dragging to the rows or columns called ***shelves***, the data source to which the fields are linked becomes the primary data source.

The primary data source is treated as authoritative for the visualization. All data for fields from the primary data source are displayed in the visualization, but only data that corresponds to those fields from secondary data sources is brought into the ***Viz***. Essentially, data from secondary sources is used to augment data from the primary source.

For example, if you linked the products database to the sales database to see all sales by product, if you used the products database as primary, all products would be listed. This includes products that have zero sales. If you used the sales database as primary, only products for which at least one sale had been made would be listed in the Viz because products with no sales wouldn't have any entries in the sales database. In the scenario it's likely a better choice to use the products database as primary so you can have a complete view of all products, including those without any sales. If you used the sales database as primary, products without sales would simply be excluded from the Viz.

**Note:** There is another issue that must be addressed when blending data where field members are not present in both data sources. In the previous example, if the products database is the primary data source, products with no sales would have null values, not zero. This issue and how to resolve it will be discussed in the next topic.

It's important to note that, while the data source connections are persistent at the workbook level, primary and secondary designations exist at the ***sheet*** level only. For each worksheet, you have to configure which data source is primary. You cannot globally identify one of the data sources as primary for all new sheets. You can copy a worksheet to create a new worksheet that preserves the primary and secondary designations, but for any new worksheets created, you will have to specify which data source is primary.

Once configured, there is no way to swap primary and secondary data sources on a worksheet. You can create a new worksheet and change which data source you add to the visualization first, thereby changing the primary data source.

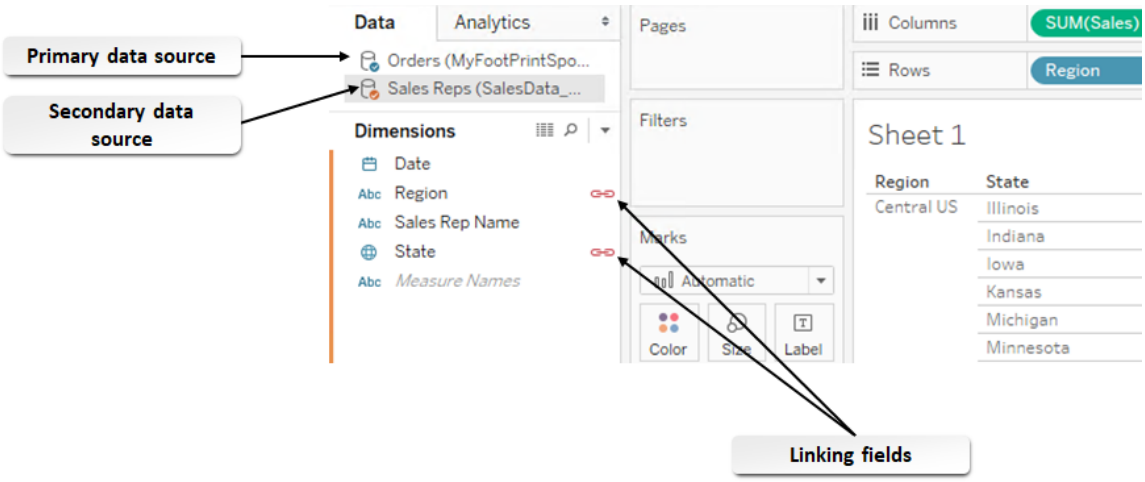



*Figure 1-4: Primary and secondary data sources.*

## Visual Indicators in Worksheets Using Data Blending

Tableau uses the following visual indicators in the user interface to indicate primary and secondary data sources as well as linking fields.

| *Visual Indicator* | *Description* |
|---|---|
| | The primary data source in the **Data** pane will have a blue check next to the data source cylinder. |
| | The secondary data sources in the **Data** pane will have an orange check next to the data source cylinder. |
| | Pill added to shelves from secondary data sources have an orange check. |
| | Linking fields in the visualization that Tableau is using to blend the data have a **Red Link** icon. Linking fields that are added to the visualization are made active. |
| | Fields that have relationships and may be used as fields to blend the data have a **Brown Link** icon. Double-click the **Broken Link** icon to make Tableau actively use the field to blend data in the visualization. You can have multiple linking fields between data sources. Adding linking fields allows you to combine more data in the Viz from secondary data sources. |

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Blend Data.** |

# ACTIVITY 1-1
## Blending Data

### Data Files

C:\095205Data\Blending Data to Visualize Relationships\MyFootPrintSports_Orders_2018.xlsx

C:\095205Data\Blending Data to Visualize Relationships\MyFootPrintSports_Sales Reps_2018.xlsx

### Before You Begin

Tableau Desktop is installed and the data files are located in the C:\095205Data folder.

### Scenario

You are a data analyst at My Footprint Sports and you use Tableau Desktop to analyze and visualize the company sales data. You've been using Tableau for a while now and are ready to expand your knowledge and what you can do. You will explore new techniques for connecting to data sources, refining and enhancing visualizations, using calculations, and forecasting data. The scenario for this course is set at the beginning of 2018. My Footprint Sports has complete data for 2014 through 2017.

You have two data sources that have a field in common but one contains records per individual sales orders and the other is a summary of the number of sales in a month per salesperson. Because the levels of granularity are different, you don't want to combine them into a single source, but you do want to combine the data for a view that will show the number of sales from one source with the associated sales amounts from the other source. You want to see as much data as you can, so you will select which data source to use as the primary based on which one gives you the most data.

**Note:** Activities may vary slightly if the software vendor has issued digital updates. Your instructor will notify you of any changes.

1. Review the Microsoft® Excel® source file.
   a) Open File Explorer and navigate to **C:\095205Data\Blending Data to Visualize Relationships**.
   b) Open **MyFootPrintSports_Orders_2018.xlsx**.

c) In Excel, review the data. This worksheet contains four years' worth of sales data. Each row is a single order.

| ID | Order ID | Order Date | Ship Date | Shipping Method | Customer ID | Customer Name | City | State | Zip Code | Country | Re |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12037 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Ce |
| 12038 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Ce |
| 12039 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Ce |
| 12040 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Ce |
| 12041 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Ce |
| 12045 | 2014-40923 | 1/15/2014 | 1/16/2014 | First Class | S-115901 | Albert Schmidt | Louisville | Kentucky | 40214 | United States | Sc |
| 12051 | 2014-40927 | 1/19/2014 | 1/22/2014 | 2 Day | W-131951 | Denise Ward | Oakland | California | 94601 | United States | W |
| 12076 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 | Aaron Anderson | Lakewood | Ohio | 44107 | United States | Ea |
| 12077 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 | Aaron Anderson | Lakewood | Ohio | 44107 | United States | Ea |
| 12098 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 | Frances Burton | Austin | Texas | 78745 | United States | Ce |
| 12099 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 | Frances Burton | Austin | Texas | 78745 | United States | Ce |
| 12105 | 2014-40947 | 2/8/2014 | 2/11/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Ce |
| 12108 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | W |
| 12109 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | W |
| 12110 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | W |
| 12111 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | W |
| 12112 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | W |
| 12113 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | W |
| 12126 | 2014-40958 | 2/19/2014 | 2/25/2014 | Standard | B-100151 | Donald Bailey | Oklahoma City | Oklahoma | 73120 | United States | Ce |
| 12130 | 2014-40960 | 2/21/2014 | 2/25/2014 | Standard | R-207401 | Alan Romero | Bloomington | Indiana | 47401 | United States | Ce |
| 12152 | 2014-40970 | 3/2/2014 | 3/7/2014 | Standard | P-107201 | Roy Payne | Florence | Alabama | 35630 | United States | Sc |
| 12153 | 2014-40970 | 3/2/2014 | 3/7/2014 | Standard | P-107201 | Roy Payne | Florence | Alabama | 35630 | United States | Sc |
| 12154 | 2014-40970 | 3/2/2014 | 3/7/2014 | Standard | P-107201 | Roy Payne | Florence | Alabama | 35630 | United States | Sc |
| 12163 | 2014-40972 | 3/4/2014 | 3/7/2014 | 2 Day | B-101501 | Chritopher Barnes | Long Beach | New York | 11561 | United States | Ea |
| 12164 | 2014-40972 | 3/4/2014 | 3/7/2014 | 2 Day | B-101501 | Chritopher Barnes | Long Beach | New York | 11561 | United States | Ea |
| 12165 | 2014-40972 | 3/4/2014 | 3/7/2014 | 2 Day | B-101501 | Chritopher Barnes | Long Beach | New York | 11561 | United States | Ea |
| 12166 | 2014-40972 | 3/4/2014 | 3/7/2014 | 2 Day | B-101501 | Chritopher Barnes | Long Beach | New York | 11561 | United States | Ea |
| 12167 | 2014-40972 | 3/4/2014 | 3/7/2014 | 2 Day | B-101501 | Chritopher Barnes | Long Beach | New York | 11561 | United States | Ea |

Orders

d) In File Explorer, in the **C:\095205Data\Blending Data to Visualize Relationships** folder, open **MyFootPrintSports_SalesReps_2018.xlsx**.

e) In Excel, review the data. This worksheet contains the different sales reps and the number of sales they had for each month in 2017.

| Date | Region | State | Sales Rep Name | Number of Sales |
|---|---|---|---|---|
| 1/31/2017 | Central US | Oklahoma | Randy Wilson | 24 |
| 1/31/2017 | Central US | Texas | Randy Wilson | 31 |
| 1/31/2017 | Eastern US | New Jersey | Larry Barnes | 24 |
| 1/31/2017 | Eastern US | New York | Larry Barnes | 34 |
| 1/31/2017 | Eastern US | Ohio | Sharon Richards | 30 |
| 1/31/2017 | Southern US | Florida | Tina Peterson | 32 |
| 1/31/2017 | Southern US | Georgia | Tina Peterson | 17 |
| 1/31/2017 | Western US | Arizona | Jean Thomas | 20 |
| 1/31/2017 | Western US | California | Jean Thomas | 36 |
| 1/31/2017 | Western US | Oregon | Timothy Adams | 28 |
| 2/28/2017 | Central US | Oklahoma | Randy Wilson | 22 |
| 2/28/2017 | Central US | Texas | Randy Wilson | 28 |
| 2/28/2017 | Eastern US | New Jersey | Larry Barnes | 26 |
| 2/28/2017 | Eastern US | New York | Larry Barnes | 33 |
| 2/28/2017 | Eastern US | Ohio | Sharon Richards | 33 |
| 2/28/2017 | Southern US | Florida | Tina Peterson | 33 |
| 2/28/2017 | Southern US | Georgia | Tina Peterson | 26 |
| 2/28/2017 | Western US | Arizona | Jean Thomas | 19 |
| 2/28/2017 | Western US | California | Jean Thomas | 30 |
| 2/28/2017 | Western US | Oregon | Timothy Adams | 31 |
| 3/31/2017 | Central US | Oklahoma | Randy Wilson | 26 |
| 3/31/2017 | Central US | Texas | Randy Wilson | 33 |
| 3/31/2017 | Eastern US | New Jersey | Larry Barnes | 24 |
| 3/31/2017 | Eastern US | New York | Larry Barnes | 33 |
| 3/31/2017 | Eastern US | Ohio | Sharon Richards | 32 |
| 3/31/2017 | Southern US | Florida | Tina Peterson | 31 |
| 3/31/2017 | Southern US | Georgia | Tina Peterson | 22 |
| 3/31/2017 | Western US | Arizona | Jean Thomas | 23 |

Sales Reps

f) Close all open **Excel** windows.

g) If prompted to save changes, select **Don't Save**.

2. Connect to the Excel spreadsheet.

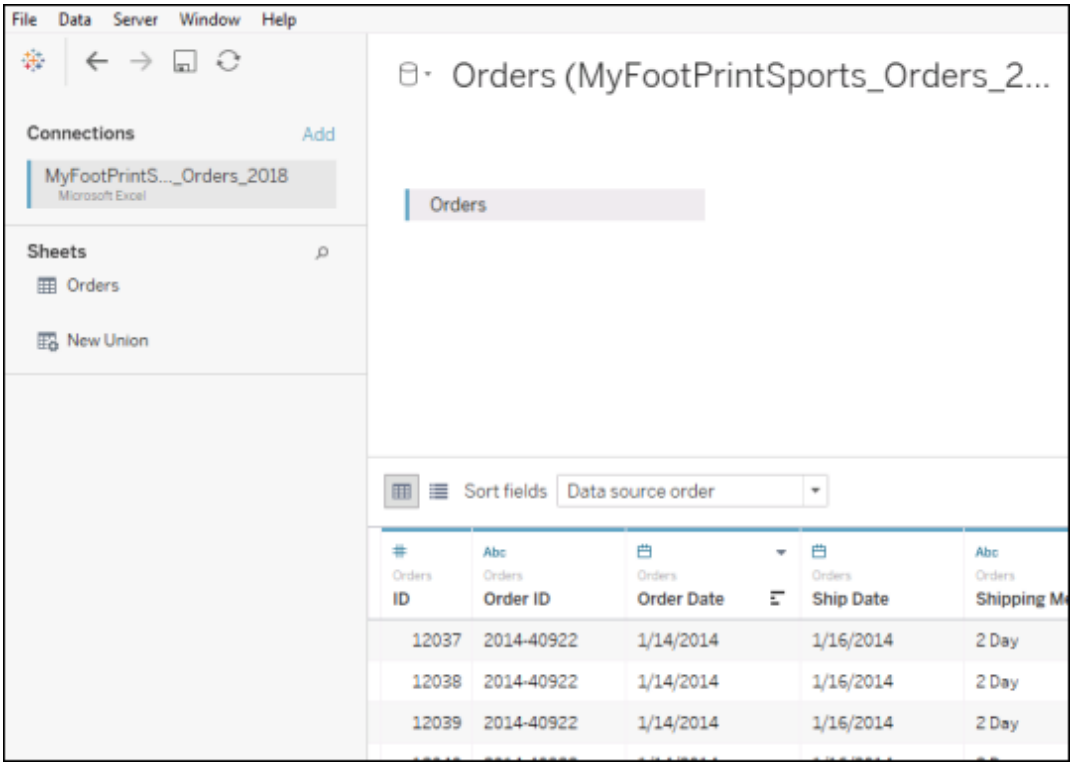
a) Open **Tableau**.

> **Note:** The trial version of Tableau Desktop expires after 14 days.

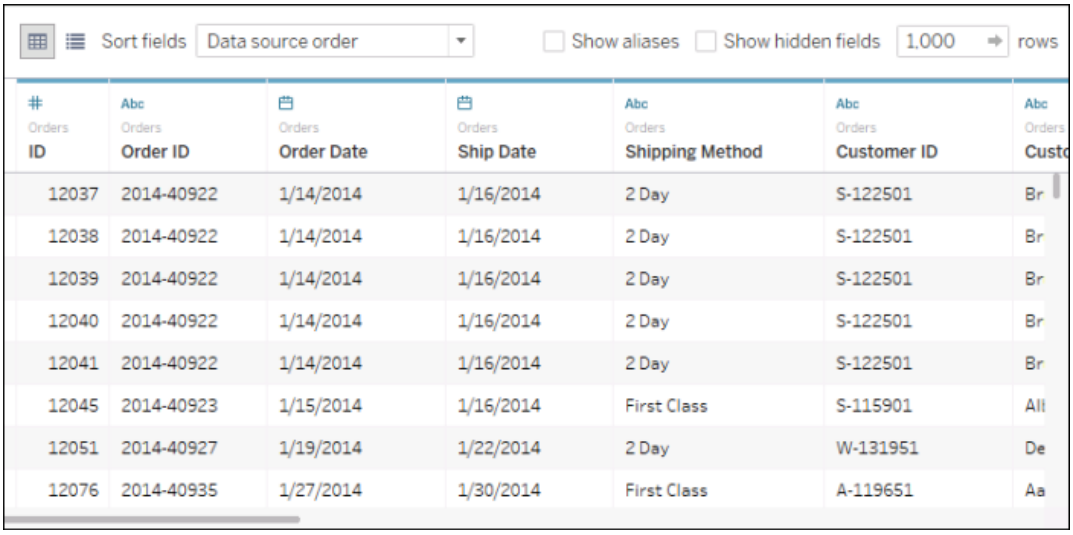
b) In Tableau, in the **Connect** pane, under **To a File**, select **Microsoft Excel**.

c) In the **Open** dialog box, navigate to **C:\095205Data\Blending Data to Visualize Relationships**.

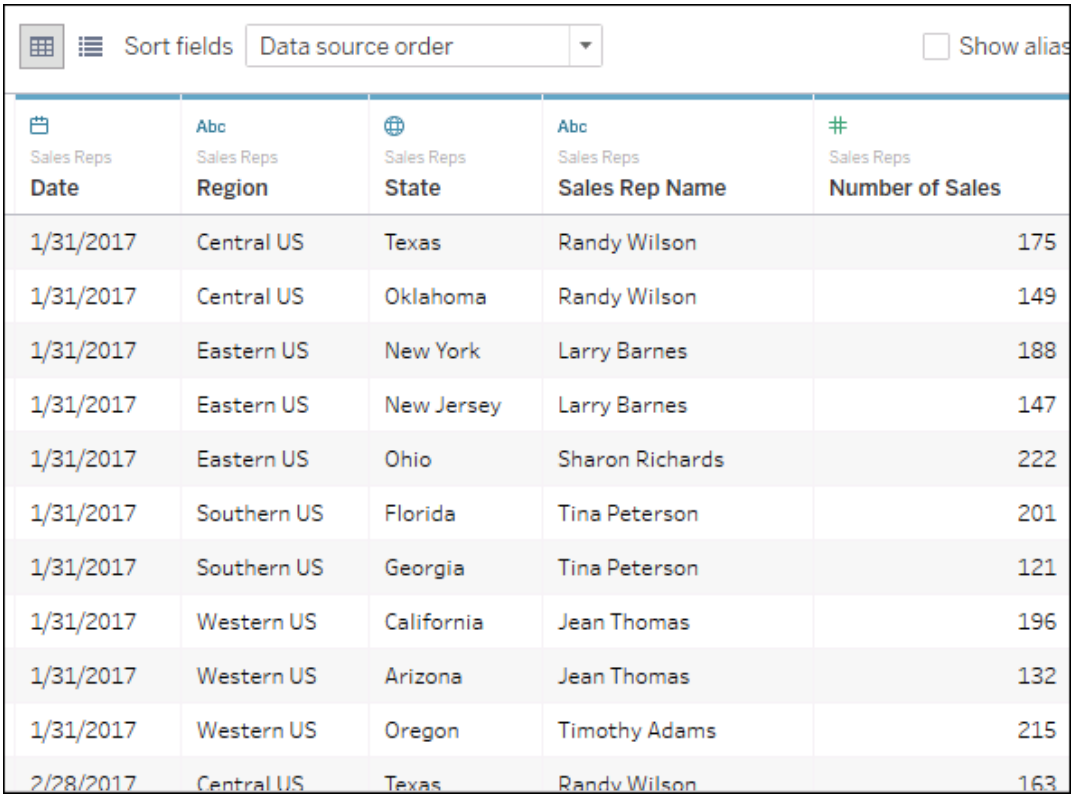
d) Select **MyFootPrintSports_Orders_2018.xlsx** and then select **Open**.
The **Data Source** page is displayed and the **Orders** sheet is automatically added to the canvas.




e) In the grid, review the data rows and columns.

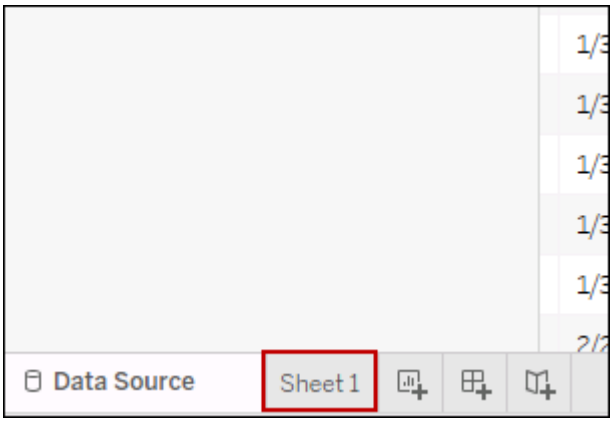


3. Create a second data connection.
   a) From the menu, select **Data→New Data Source**.
   b) Under **To a File**, select **Microsoft Excel**.
   c) In the **Open** dialog box, select **MyFootPrintSports_SalesReps_2018.xlsx** and then select **Open**.
   d) In the grid, review the data rows and columns.

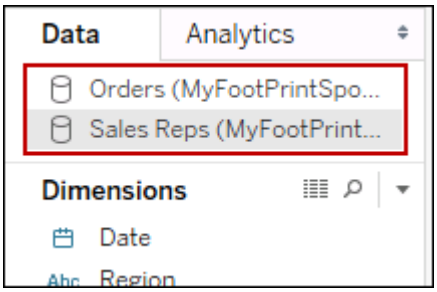



| Sales Reps Date | Sales Reps Region | Sales Reps State | Sales Reps Sales Rep Name | Sales Reps Number of Sales |
|---|---|---|---|---|
| 1/31/2017 | Central US | Texas | Randy Wilson | 175 |
| 1/31/2017 | Central US | Oklahoma | Randy Wilson | 149 |
| 1/31/2017 | Eastern US | New York | Larry Barnes | 188 |
| 1/31/2017 | Eastern US | New Jersey | Larry Barnes | 147 |
| 1/31/2017 | Eastern US | Ohio | Sharon Richards | 222 |
| 1/31/2017 | Southern US | Florida | Tina Peterson | 201 |
| 1/31/2017 | Southern US | Georgia | Tina Peterson | 121 |
| 1/31/2017 | Western US | California | Jean Thomas | 196 |
| 1/31/2017 | Western US | Arizona | Jean Thomas | 132 |
| 1/31/2017 | Western US | Oregon | Timothy Adams | 215 |
| 2/28/2017 | Central US | Texas | Randy Wilson | 163 |

4. Review connected data sources.
   a) At the bottom of the page, select the **Sheet 1** worksheet.



b) In the **Data** pane, observe that both data sources are listed.



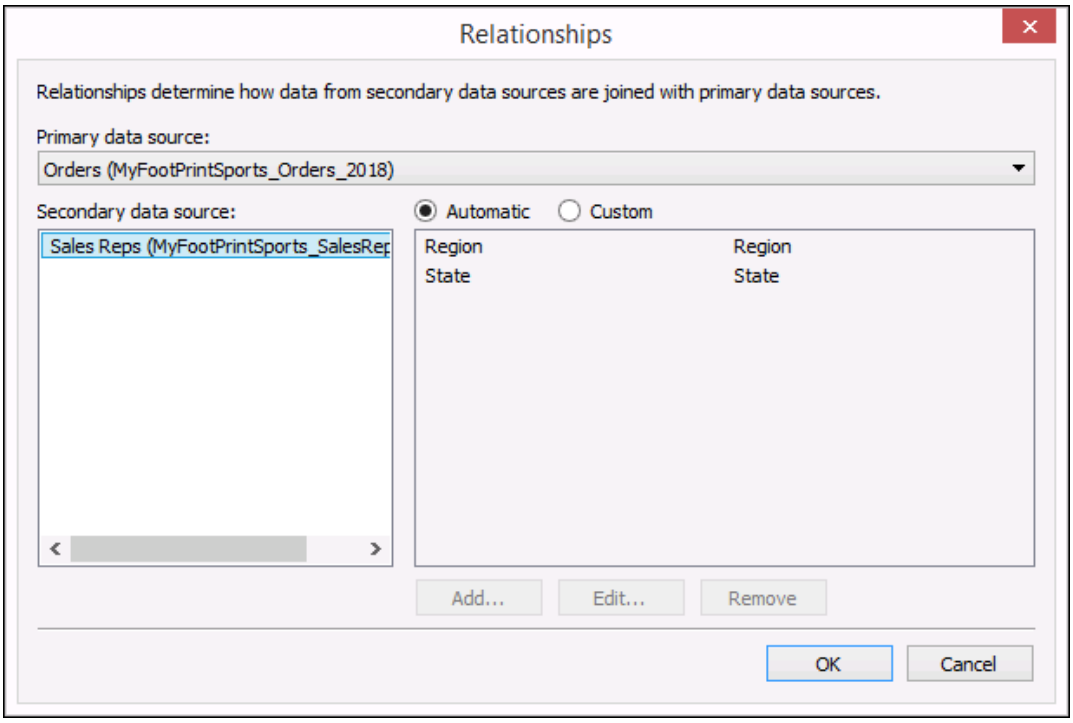

c) In the **Data** pane, review the **Dimensions** and **Measures** for the **Sales Reps** data source. Note the **Region** and **State** fields.

d) In the **Data** pane, select the **Orders** data source.

e) In the **Data** pane, review the **Dimensions** and **Measures** for the **Orders** data source. Note the **Region** and **State** fields are present for this data source as well.

f) From the menu, select **Data→Edit Relationships**.

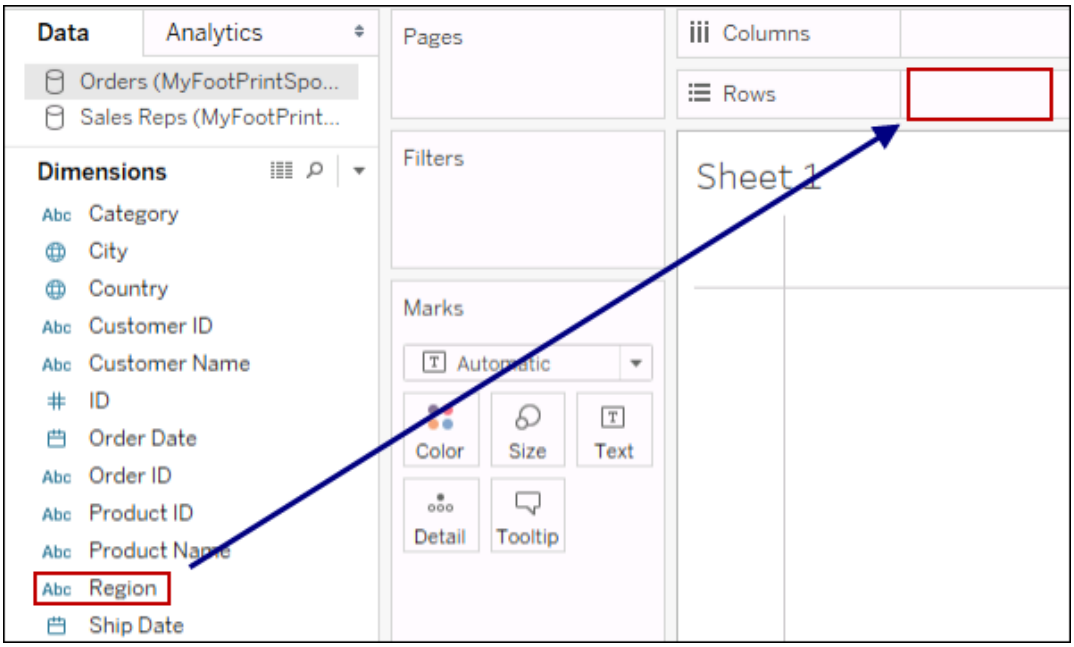
g) In the **Relationships** dialog box, observe that Tableau automatically created a relationship between the two data sources using the **Region** and **State** fields.




h) Select **Cancel** to close the dialog box without making a change.
You only wanted to review the relationships that Tableau made automatically.

5. Create a view using the **Orders** data source.

a) In the **Data** pane, under **Dimensions**, select and drag **Region** to the **Rows** shelf.



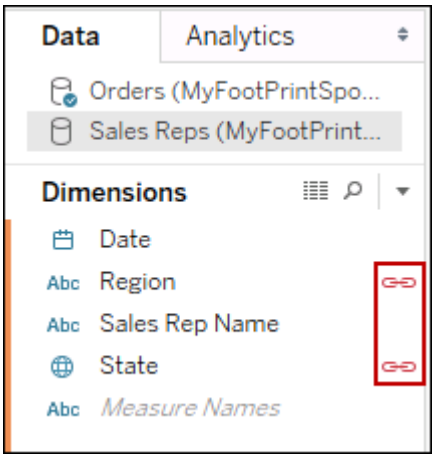

b) In the **Data** pane, under **Dimensions**, select and drag **State** to the **Rows** shelf to the right of **Region**.

c) In the **Data** pane, under **Measures**, select and drag **Sales** to the **Columns** shelf.

## 6. Add fields from the **Sales Reps** data source.

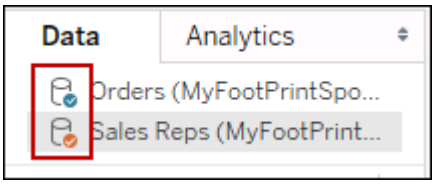
a) In the **Data** pane, select the **Sales Reps** data source.

b) In the **Data** pane, under **Dimensions**, observe the orange link icon for **Region** and **State**. This **Red Link** icon notifies you that these fields are linked to fields in the primary data source.

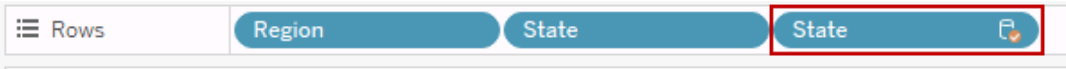


c) In the **Data** pane, under **Dimensions**, select and drag **State** to the **Rows** shelf to the right of the existing **State** pill.

In the **Data** pane, observe the blue checkmark for the **Orders** data source, and the orange checkmark for the **Sales Reps** data source. Blue marks the data source as primary and orange marks it as secondary.

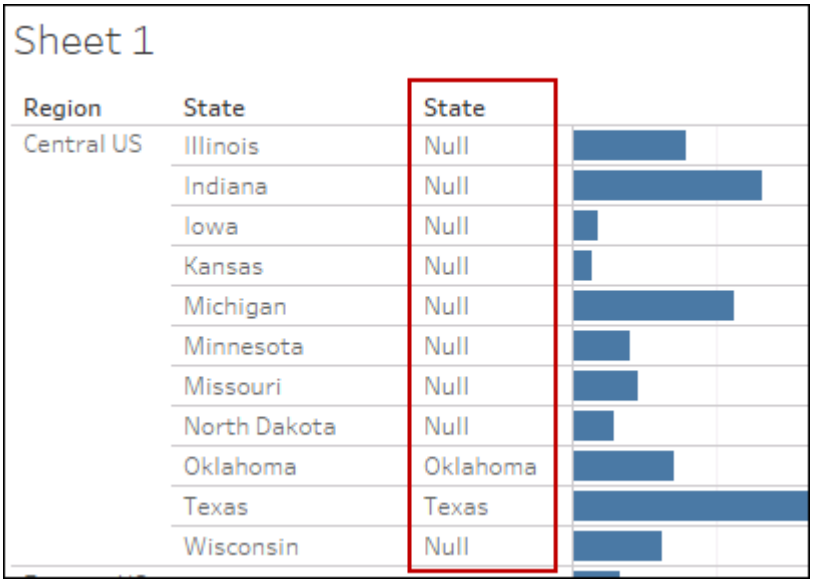



In the **Rows** shelf, observe the same orange checkmark data source icon for the second **State** pill.




7. Add **Number of Sales** to the view.

a) In the canvas, observe the null values for many of the values in the second **State** field. This is because the secondary data source does not have values for as many of the states as the primary data source.




b) In the **Data** pane, under **Measures**, select and drag **Number of Sales** to the **Columns** shelf to the right of the **SUM(Sales)** pill.

c) In the canvas, observe the values for the **Number of Sales** field and how there is only data for some of the states.

8. Rename the **Sheet 1** worksheet.

a) Right-click the **Sheet 1** tab and select **Rename**.

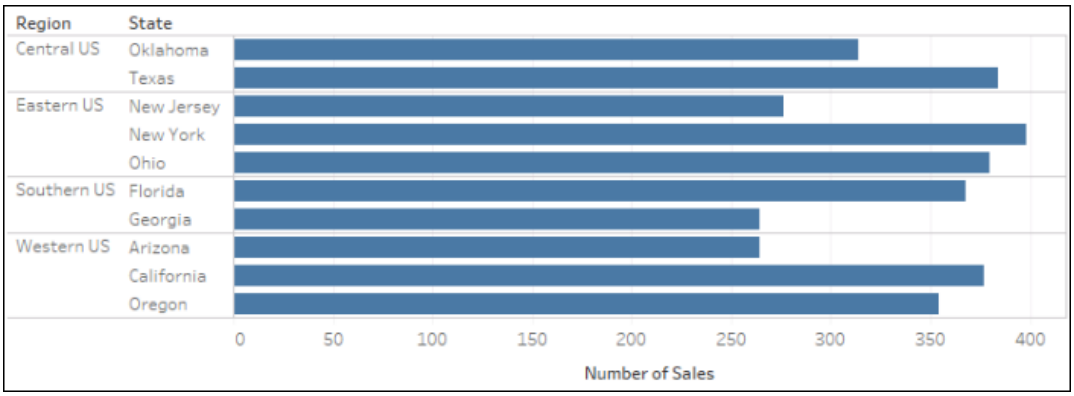
b) Type ***Orders Primary*** and press **Enter**.

9. Create a view using **Sales Reps** as the primary data source.

a) On the toolbar, select the **New Worksheet** button.

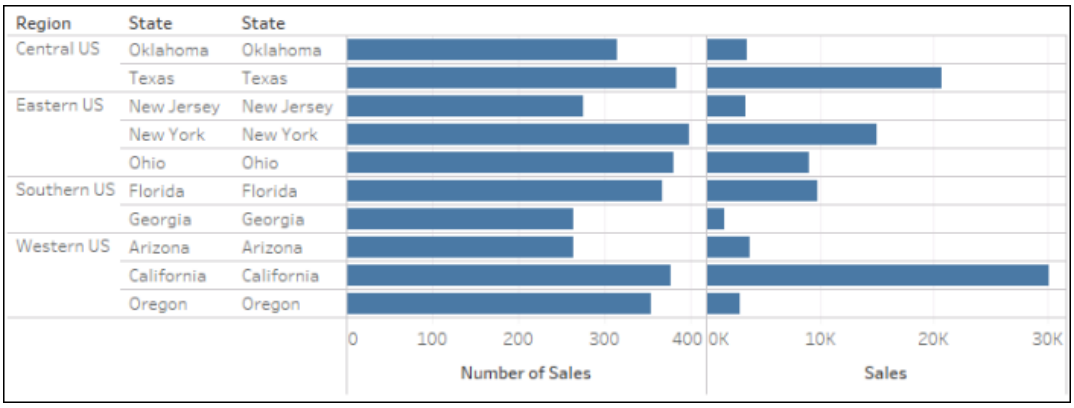
b) On the **Sheet 2** tab, in the **Data** pane, verify the **Sales Reps** data source is selected.

c) In the **Data** pane, under **Dimensions**, select and drag **Region** to the **Rows** shelf.
d) In the **Data** pane, under **Dimensions**, select and drag **State** to the **Rows** shelf to the right of **Region**.
e) In the **Data** pane, under **Measures**, select and drag **Number of Sales** to the **Columns** shelf.
f) In the canvas, observe the number of states that are listed from the **Sales Reps** data source.




## 10. Add fields from the **Orders** data source to the view.

a) In the **Data** pane, select the **Orders** data source.
b) In the **Data** pane, under **Dimensions**, observe the orange link icon for **Region** and **State**.
c) In the **Data** pane, under **Dimensions**, select and drag **State** to the **Rows** shelf to the right of the existing **State** pill.
d) In the **Data** pane, under **Measures**, select and drag **Sales** to the **Columns** shelf to the right of the **SUM(Number of Sales)** pill.
e) In the canvas, observe the number of states did not change after adding fields from the **Orders** data source.




## 11. Rename the **Sheet 2** worksheet.

a) Right-click the **Sheet 2** tab and select **Rename**.
b) Type ***Sales Reps Primary*** and press **Enter**.

## 12. Compare the Orders Primary and Sales Reps Primary worksheets. How are the two views different and why?

13. Which data source used as the primary achieves your goal of displaying as much data as possible?

_____

# TOPIC B

## Troubleshoot and Refine Data Blends

Blending data may result in data anomalies and other unforeseen data issues. In this topic, you will troubleshoot and refine data blends.

### Null Values

In a data source, a ***null value*** means there is no data. This is an important concept to understand. A null value means the absence of data or nothingness. If the field in question is a numeric field, a null value does not mean the value is zero. The value is null. The same goes for string values. Most databases, applications, and ***data visualization*** software have strict requirements about how the application can deal null values. In most cases, null values must be addressed at some level for calculations, aggregations, or other activities—such as combining strings—to function properly. For example, in Tableau, in a calculation where a null value is calculated with a real number, the result is always a null value.

There are many mechanisms used at many levels to address issues around null values. For example, a data-driven application may insert a default value for fields in a form rather than leaving them empty, which would mean the record contained a null value. In many cases, null values are converted into something that can be processed during data preparation.

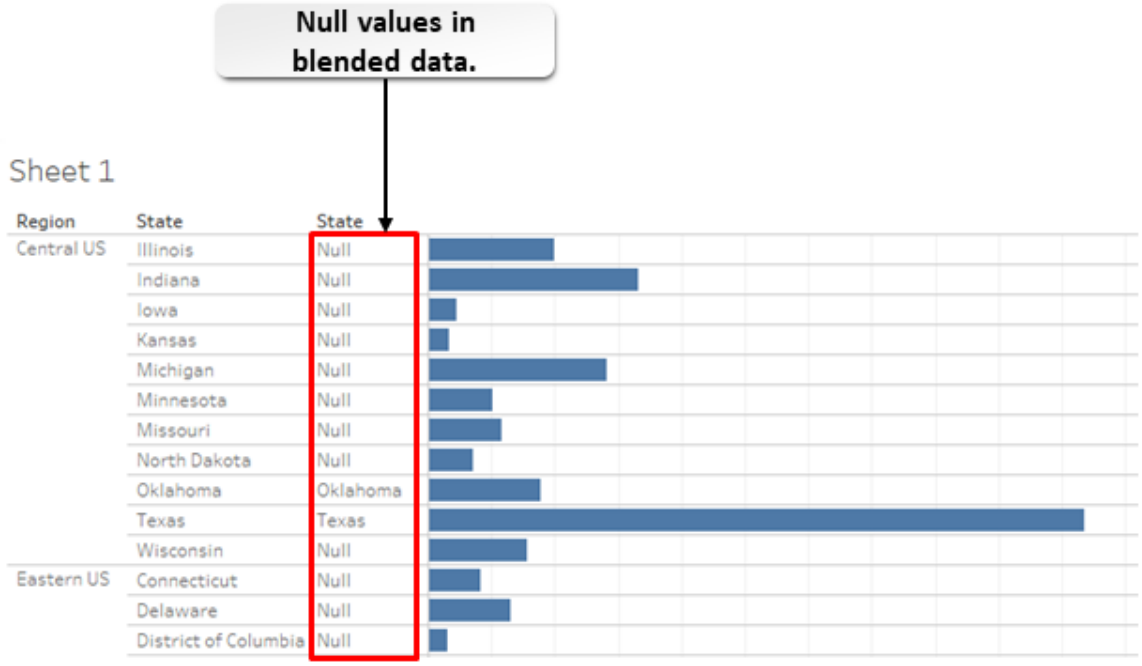



*Figure 1-5: Null values.*

### Null Values in Blended Data

Null values are a concern in data blending because, when mixing data from two sources, it's possible data from secondary sources may not exist for all records available in the primary data source. The following table shows what can cause null values in the visualization and shows some options for addressing them.

| Issue | Example | Potential Resolution |
|---|---|---|
| A secondary data source has no data for a linked field in the primary. | A calculated field that combines sales data from two data sources not showing data from the primary data source. | Calculations between nulls and real numbers generate nulls so you have to replace null values. You can update the secondary data source to:<br>• Insert values in place of null values as part of an update to the underlying data source.<br>• Replace null as part of data preparation.<br>• Use the ZN function in a calculation to replace nulls with zeros. |
| Linked field data types are different. | A linked field is type Number in one data source and String in another. | Verify linked fields are the same type. If they are not, create a new field in one of the data sources that matches the field type in the other, and populate it with data from the non-matching field. You may have to massage the original data to obtain the format you need for the new field. |

*Figure 1-6: Null values in blended data.*

### ZN Function

You can use the ZN (Zero Null) function in Tableau as part of the calculation to convert any nulls from secondary data sources to zeros.

## Unexpected Null Values in Data

To avoid unexpected values in visualizations, the following characteristics should be identical between linked fields in primary and secondary data sources:

- Field length
- Case
- Values

If data characteristics don't match, Tableau will attempt to reconcile the issues and convert the data to identical values based on those in the primary data source. This may have unexpected results.

For example, if a primary and secondary data source are linked on the State field, and the primary data source contains the value "oregon" while the secondary data source contains the value "Oregon," the value displayed will be "oregon."

If values cannot be successfully reconciled due to field length issues, misspellings, or other circumstances, null values may be the result.

### Additional Information

For more information, see **https://kb.tableau.com/articles/issue/Nulls-Show-When-Data-Blending**.

## Asterisks in Blended Visualizations

When blending data and selecting linking fields, you should try to ensure secondary data sources only have one value for each record in the primary data source. If a secondary data source has more than one value for a mark from the primary data source, an asterisk will be displayed in the visualization. For example, if you wanted to view product sales by ZIP Code, you could create a data blend using your sales database as primary and your customer database that contains ZIP Code information as secondary. If you linked the two on the State field, any state that has more than one ZIP Code will return an asterisk because there are multiple values; in this case, multiple ZIP Codes, for most states.

In some cases, you can resolve this by creating a new worksheet and switching which data source is primary and which is secondary. It's a general best practice to use the data source that has more granular information as the primary data source. You can also opt to use a higher level linking field. For example, you might link the two data sources on the StoreID field to see sales by ZIP Code for each store.

## Common Warnings and Errors

The following are common warnings and errors, and potential causes of these you might see when blending data in Tableau.

| Error | Cause | Potential Action |
|---|---|---|
| Fields cannot be used from the [Secondary] data source, because there is no relationship to the primary data source. | There are not active linking fields in the secondary data source or there are no relationships established. | Create relationships between the primary and secondary data sources if need be. In the **Data** pane, click the secondary data source and find the inactive linking field icon for the fields you wish to blend on and double-click to make the link active. |
| The primary and secondary connections are from tables in the same data source. | The workbook has separate data sources connected to the same database. | Create a join in the underlying database rather than using data blending in Tableau. Joins can integrate data from many tables and may improve performance and filtering. |
| Cannot blend the secondary data source because one or more fields use an unsupported aggregation. | Results from non-additive aggregates including COUNTD, MEDIAN, and RAWSQLAGG cannot be aggregated along dimensions. | Use different aggregations or create calculated fields to show results. |

*Figure 1-7: Common warnings and errors.*

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Refine Blends to Visualize Key Information.**

# ACTIVITY 1-2

## Refining Blends to Visualize Key Information

### Before You Begin

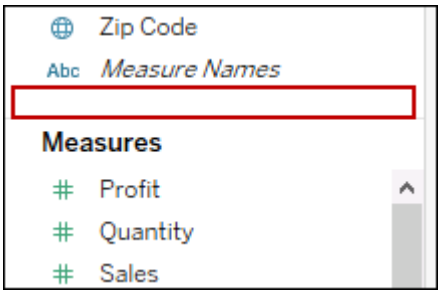
Tableau Desktop is open.

### Scenario

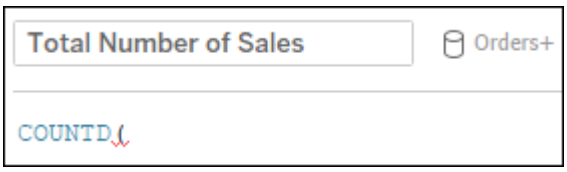
You already have the number of sales from the sales reps, but you also want to find the number of sales that came from customers purchasing directly from stores. You want to take this number of sales from direct customers and add it with the sales reps number of sales to get a total number of sales. Then you will compare those numbers in your view.

---

1. Create calculated field to add the number of sales from sales reps to the number of unique orders from the **Orders** data source.
   a) At the bottom of the page, select the **Orders Primary** worksheet.
   b) In the **Data** pane, verify the **Orders** data source is selected.
   c) In the **Data** pane, under **Dimensions** or **Measures**, right-click in empty white space and select **Create Calculated Field**.




   d) In the text box, type ***Total Number of Sales***
   e) In the code area, type `COUNTD(`. The **COUNTD** function will count the number of unique order IDs in the **Orders** data source.




| | |
|---|---|
| | **Note:** You can refer to the images and use those as a guide when typing calculation code in this course. |
| | **Note:** As you type, the autofill will display options for what you are typing. |

   f) Position the **Total Number of Sales** dialog box so you can see it and the **Data** pane at the same time.
   g) In the **Data** pane, under **Dimensions**, select and drag **Order ID** to the **Total Number of Sales** dialog box to the end of the line.

h) In the code area, at the end of the line of code, type )

```
COUNTD([Order ID])

The calculation is valid.
```

| | |
|---|---|
| | **Note:** At the bottom of the code area, Tableau will indicate if the calculation code is valid or invalid. If your code is invalid, the first places to check are ensuring brackets and parentheses are in the correct place. |

## 2. Finish the calculated field.

a) In the code area, at the end of the line of code, type +

b) In the **Data** pane, select the **Sales Reps** data source.

c) In the **Data** pane, under **Measures**, select and drag **Number of Sales** to the code area, at the end of the line of code.

d) In the code area, at the end of the line of code, verify that there is a closing )

e) Select **OK**.

## 3. Add calculated field to view.

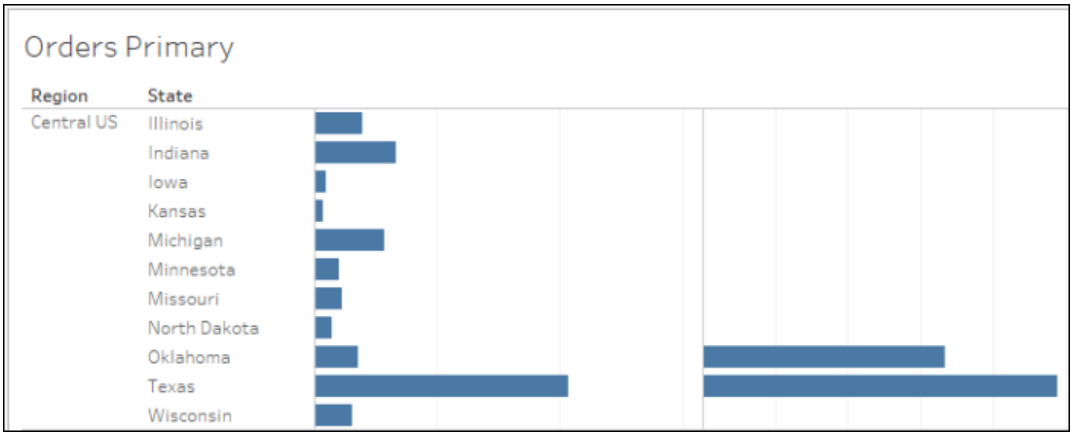
a) In the **Columns** shelf, right-click **SUM(Number of Sales)** and select **Remove**.

b) In the **Rows** shelf, right-click the second **State** pill and select **Remove**.

c) In the **Data** pane, verify that the **Orders** data source is selected.

d) In the **Data** pane, under **Measures**, select and drag **Total Number of Sales** to the **Columns** shelf.

e) Observe that the number of sales for states that have data in the **Sales Reps** data source is showing. This is because the values for the states that are not in the **Sales Reps** data source are null and the calculation needs a number such as a zero to function correctly.




## 4. Convert null values to zero in calculated fields.

a) In the **Data** pane, under **Measures**, right-click **Total Number of Sales** and select **Edit**.

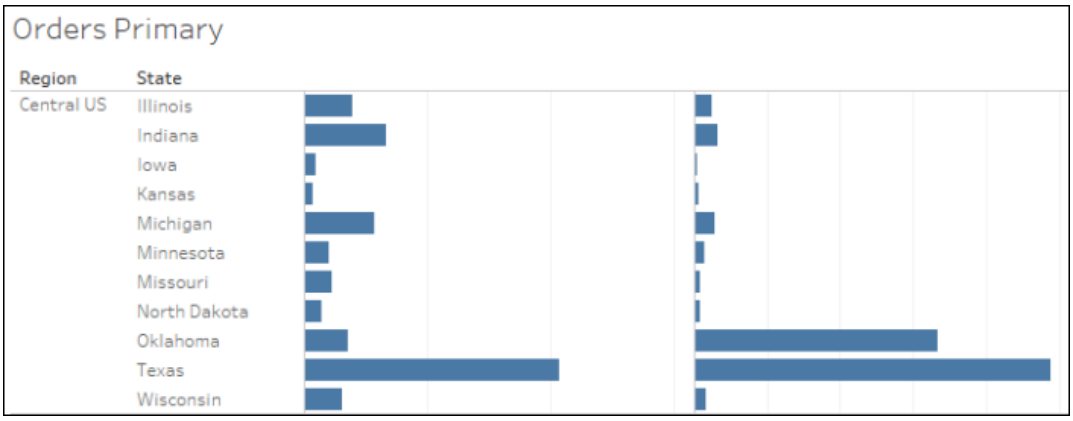
b) In the code area, place the cursor to the right of the plus sign. You will use the ZN function to convert null values to zeros.

c) Type `ZN(`
The ZN function checks for a null value, and if the value is null, the function changes it to zero.

d) Place your cursor at the end of the calculation and type `)`

e) Select **OK**.

f) Observe that the other states are now showing the number of sales values.



5. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Blending Data to Visualize Relationships** folder.
   c) In the **File name** box, type ***MyCopy_MyFootPrintSports_2018_1*** and select **Save**.
   d) From the menu, select **File→Close**.

# Summary

In this lesson, you combined data from different data sources using data blending, fixed issues with blended data including null values, and worked on troubleshooting data resulting from blends.

### In your environment, when might you create views from multiple data sources? Would blending be a good choice?

### What problems can null values in blended data cause in your environment?

**Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience.

# 2 Joining Data

**Lesson Time: 2 hour, 15 minutes**

## Lesson Introduction

Another way to combine data in Tableau® is to join data. In this lesson, you will join data to combine it in Tableau, then troubleshoot joins to make sure the data is accurate, and finally use unions to merge data.

## Lesson Objectives

In this lesson, you will:

- Create joins.
- Troubleshoot joins.
- Merge data with unions.

# TOPIC A

## Create Joins

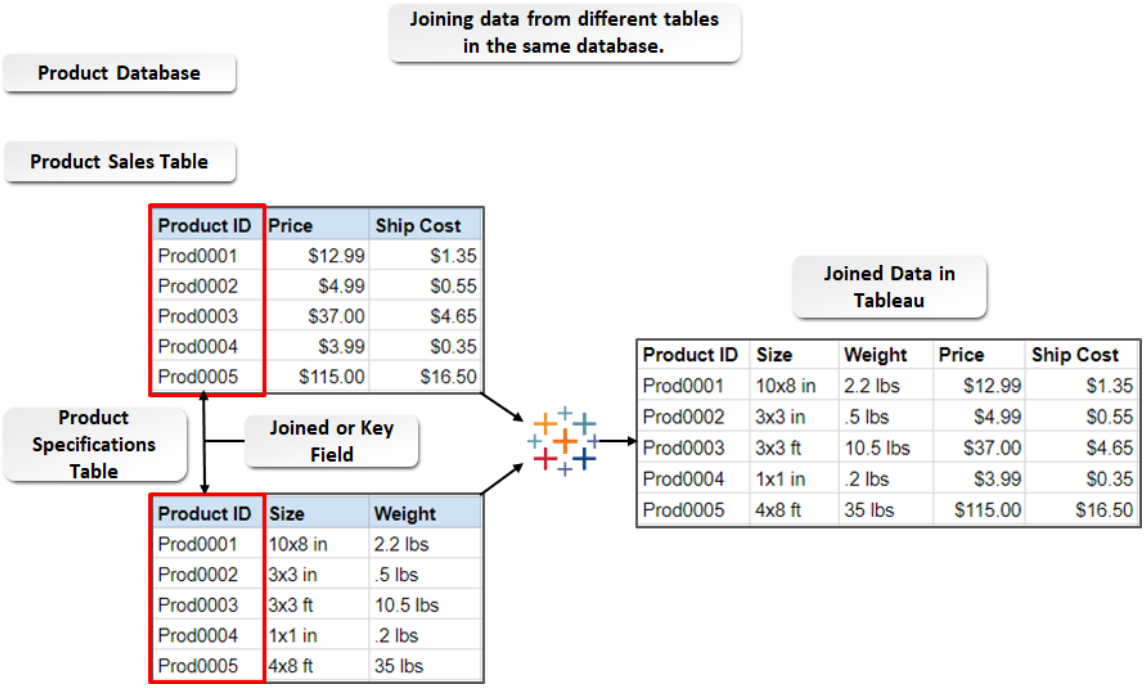
Joins are another method available in Tableau to combine data. In this topic, you will create joins to combine data for visualizations.

### Joins

The word *join* is a database term that means to combine columns from one or more tables in a relational database. The join uses a common field present in both tables to create a new table with information from each joined table. By combining the data from the two joined tables, the new table shows more columns of information about data, which allows you to perform more robust analysis and build visualizations that include more comprehensive information.

Joins can be performed at the database level to two or more separate tables, allowing users to query and retrieve information from those different tables. Tableau can also join tables in a single database, or multiple worksheets in a Microsoft® Excel® workbook, to combine data. Tableau also supports cross-database joins that can link tables in separate data sources. Joined fields must be the same data type.

For example, a Product Specifications table may have information about every product, such as size and weight. The Product Sales table stores data on costs, sales price, and shipping expenses. Both might have a field named Product ID that has each product's unique stock keeping unit (SKU) number. You might join the two tables on the Product ID field to ensure data is properly related between the tables.




| Product ID | Price | Ship Cost |
|---|---|---|
| Prod0001 | $12.99 | $1.35 |
| Prod0002 | $4.99 | $0.55 |
| Prod0003 | $37.00 | $4.65 |
| Prod0004 | $3.99 | $0.35 |
| Prod0005 | $115.00 | $16.50 |

| Product ID | Size | Weight |
|---|---|---|
| Prod0001 | 10x8 in | 2.2 lbs |
| Prod0002 | 3x3 in | .5 lbs |
| Prod0003 | 3x3 ft | 10.5 lbs |
| Prod0004 | 1x1 in | .2 lbs |
| Prod0005 | 4x8 ft | 35 lbs |

| Product ID | Size | Weight | Price | Ship Cost |
|---|---|---|---|---|
| Prod0001 | 10x8 in | 2.2 lbs | $12.99 | $1.35 |
| Prod0002 | 3x3 in | .5 lbs | $4.99 | $0.55 |
| Prod0003 | 3x3 ft | 10.5 lbs | $37.00 | $4.65 |
| Prod0004 | 1x1 in | .2 lbs | $3.99 | $0.35 |
| Prod0005 | 4x8 ft | 35 lbs | $115.00 | $16.50 |

*Figure 2-1: Joins.*

Once joined, you could then build visualizations that show sales figures and shipping costs related to each product to negotiate new shipping agreements or to adjust prices for changes in shipping rates.

Joins are made on a specific field or fields. Not all join types are supported by all data sources—such as cube or support joins—in Tableau.

**Note:** To learn more about the differences of cube data sources, check out the Spotlight on **Differences with Cube Data Sources** presentation from the **Spotlight** tile on the CHOICE Course screen.

Tableau supports joins for linking tables in the same database or linking worksheets in the same workbook, and ***cross-database joins*** for joining tables and worksheets in different data sources.

**Note:** When and why to use cross-database joins and data blending depends on many factors including the type of data connection and performance needs.

## Integrated Data Sources

When you create a cross-database join, you are creating what is called an ***integrated data source***. An integrated data source is a data source that connects to more than one external data source, such as separate databases or workbooks through a cross-database join. Functionally, a cross-database connection and an integrated data source can be thought of as the same thing.

Integrated data sources appear as a single data source connection in the **Data** ***pane***. Compare this to blends where both data sources appear in the **Data** pane and where you can change which data source is primary and which is secondary on each worksheet. Integrated data sources can also be saved and published to ***Tableau Server*** or ***Tableau online*** like other data sources, where blends cannot.

Fields are grouped by the data source from which they came by default under integrated data sources. However, you can turn this behavior off and choose to have fields grouped by folder.

## Impact of Join Types

The type of join you use determines what data Tableau, or other databases and data visualization software that use joins, will be combined. The type of join also determines what will not be combined based on how the data in the joined fields matches between the joined tables. In a perfect world, both joined tables would have identical data in the fields that are joined; however, this is often not the case.

Continuing the previous example where the Product Specifications and Product Sales tables were joined on the Product ID field, while the Product Specifications table might have all product IDs for every product sold, the Product Sales table may not. Any products that have not been sold will not have rows in that table.

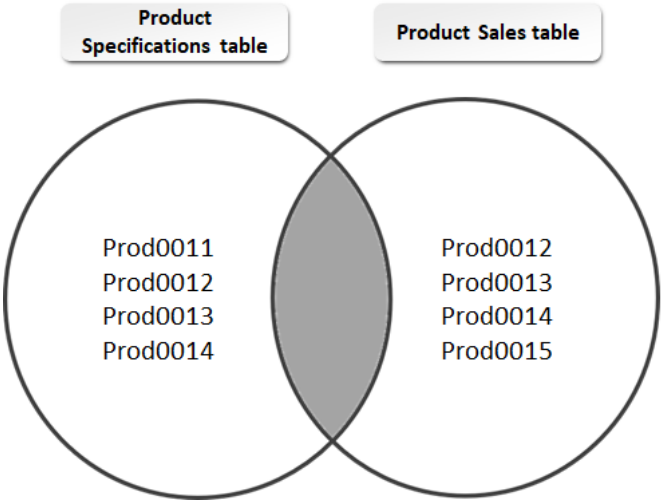
The type of join you select dictates how Tableau or other databases will deal with missing or mismatched data in the fields that are joined, and what data will be included or excluded when data is combined as a result of the join.

## Inner Joins

An inner join combines only the rows in the output that have the same data in the joined field.

For example if the Products Specifications table and the Product Sales table are joined with an inner join on the Product ID field, only rows for Prod12, Prod13, and Prod14 will be included in the combined data. The Prod11 entry does not exist in the Sales Table as that product has had no sales to date. The Prod15 entry was added for warranties that are sold with products. Since this is not technically a product but a service, it has no entry in the Product Specification table.

One benefit of inner joins is that only rows that have data are combined, so no fields will have null values.




Result

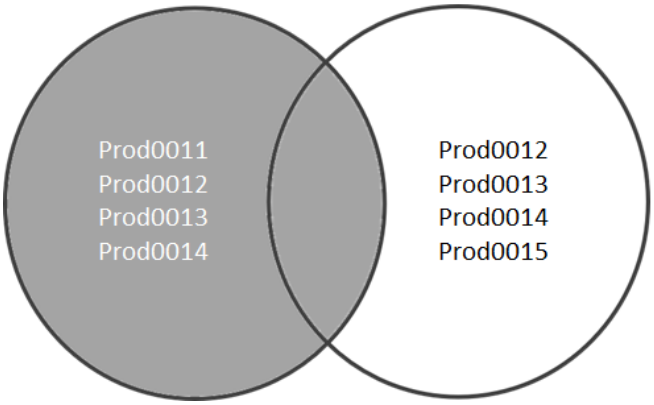
| Product ID | Weight | Price |
|---|---|---|
| Prod0012 | 1.5 lbs | $3.99 |
| Prod0013 | 10.5 lbs | $34.00 |
| Prod0014 | 2.2 lbs | $6.99 |

*Figure 2–2: Inner joins.*

## Left Joins

A left join treats the left table as the primary data source. The combined data will include all of the rows from the left table, and any related information corresponding to those rows from the right table. If there is not data available for a particular row in the right table, a null value is placed in the empty column.

For example, if the Products Specifications table and the Product Sales table are joined with a left join on the Product ID field, only rows for Prod11, Prod12, Prod13, and Prod14 will be included in the combined data as those are the Product ID rows available in the left table. Since Prod11 has no entry in the Product Sales database, the Price and Shipping Costs columns will contain null values for Prod11 in the combined table.




Result

| Product ID | Weight | Price |
|---|---|---|
| Prod0011 | 3.5 lbs | |
| Prod0012 | 1.5 lbs | $3.99 |
| Prod0013 | 10.5 lbs | $34.00 |
| Prod0014 | 2.2 lbs | $6.99 |

*Figure 2–3: Left joins.*

### Comparing Joins to Data Blends

Data blending is most similar to creating a left join. Aggregation of data occurs at different times in blending and joining. With joins, the results of the join are sent back to Tableau and aggregated. With a blend, data is aggregated first, then combined by Tableau.

## Right Joins

A right join is similar to a left join, but reverses which table is treated as primary. A right join treats the right table as the primary data source. The combined data will include all of the rows from the right table and any related information corresponding to those rows from the left table. If there is not data available for a particular row in the left table, a null value is placed in the empty column.

For example, if the Products Specifications table and the Product Sales table are joined with a right join on the Product ID field, only rows for Prod12, Prod13, Prod14 and Prod15 will be included in the combined data as those are the Product ID rows available in the right table. Since Prod15 has no entry in the Product Specifications database, the Weight and Size columns will contain null values for Prod15 in the combined table.

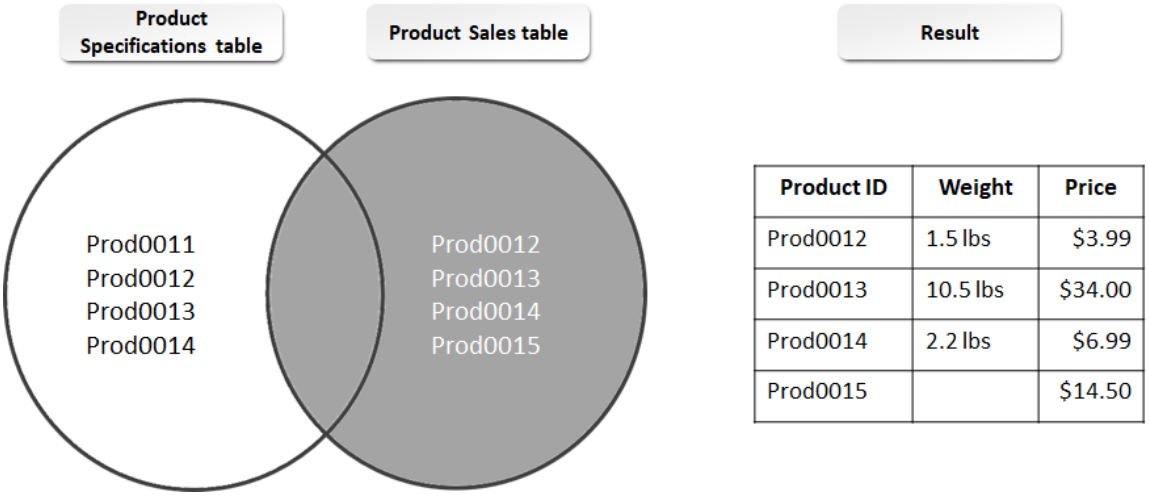
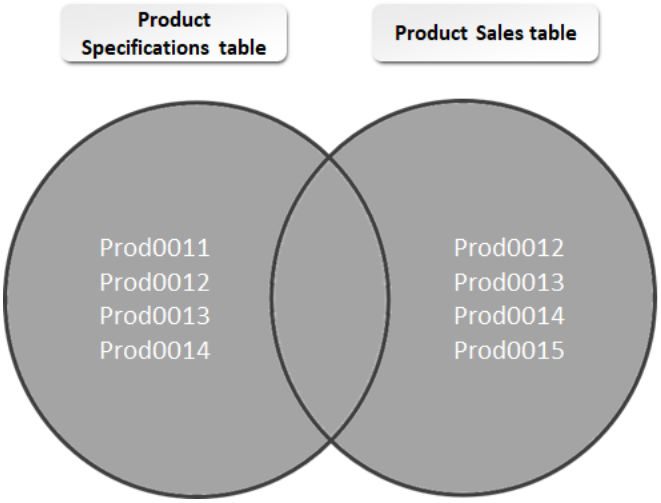



*Figure 2-4: Right joins.*

## Outer Joins

Outer joins are like the catch all of join types. An outer join combines all joined rows from both tables and any data corresponding to the rows from the other table. Null values will be entered into any columns that do not have data in the corresponding table.

For example, if the Products Specifications table and the Product Sales table are joined with an outer join on the Product ID field, the rows for Prod11, Prod12, Prod13, Prod14 and Prod15 will be included in the combined data. Since Prod11 has no entry in the Product Sales database, the Price and Shipping Costs columns will contain null values for Prod11 in the combined table. Since Prod15 has no entry in the Product Specifications database, the Weight and Size columns will contain null values for Prod15 in the combined table.




Result

| Product ID | Weight | Price |
|---|---|---|
| Prod0011 | 3.5 lbs | |
| Prod0012 | 1.5 lbs | $3.99 |
| Prod0013 | 10.5 lbs | $34.00 |
| Prod0014 | 2.2 lbs | $6.99 |
| Prod0015 | | $14.50 |

*Figure 2-5: Outer joins.*

## Join Type Selection

When you add sheets to the canvas, Tableau will join them. You can double-click the **Join** icon to configure the join. You can select the type of join you wish to use, and see the fields that Tableau joined.

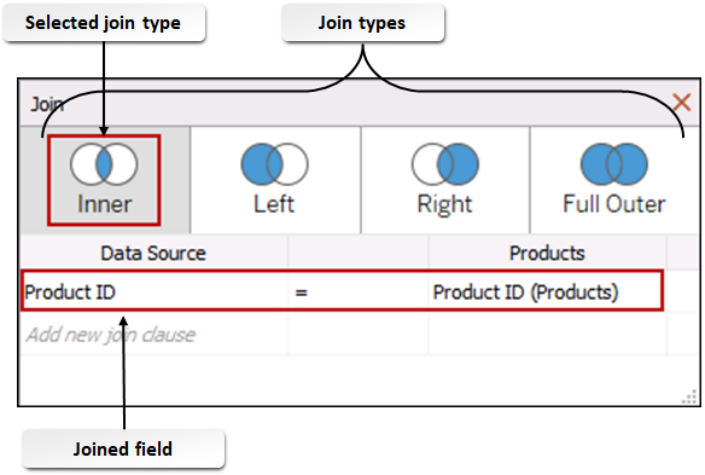



*Figure 2-6: Join type selection.*

You can also add a new join clause to join the tables based on a different field. You can also set the join condition for the fields joined. The join condition determines which related data will be brought in through the join.

The following table lists the join conditions available in Tableau.

| *Join Condition* | *Description* |
|---|---|
| = | Joined fields are equal to each other. This is the most common join condition. Join columns where the ProductID in the left table ***is the same*** as the ProductID in the right table. |
| <> | Joined fields are not equal to each other. Join columns where the ProductID in the left table ***is not the same*** as the ProductID in the right table. |

| Join Condition | Description |
|---|---|
| < | The left field value is less than the right field value. Used with numeric or date key fields. Join columns where the value in the joined field in the left table ***is less than*** the value in the right table. |
| <= | The left field value is less than or equal to the right field value. Used with numeric or date key fields. Join columns where the value in the joined field in the left table ***is less than or equal to*** the value in the right table. |
| > | The left field value is greater than the right field value. Used with numeric or date key fields. Join columns where the value in the joined field in the left table ***is greater than*** the value in the right table. |
| >= | The left field value is greater than or equal to the right field value. Used with numeric or date key fields. Join columns where the value in the joined field in the left table ***is greater than or equal to*** the value in the right table. |

## Connections and Cross-Database Joins

To create a cross-database join, you must add a new connection from the **Connections** area on the *Data Source page*. When you do this, you're telling Tableau to add a new table from a new source. Next to **Connections**, click **Add**, then select a connection type from the **Add a Connection** dialog box.

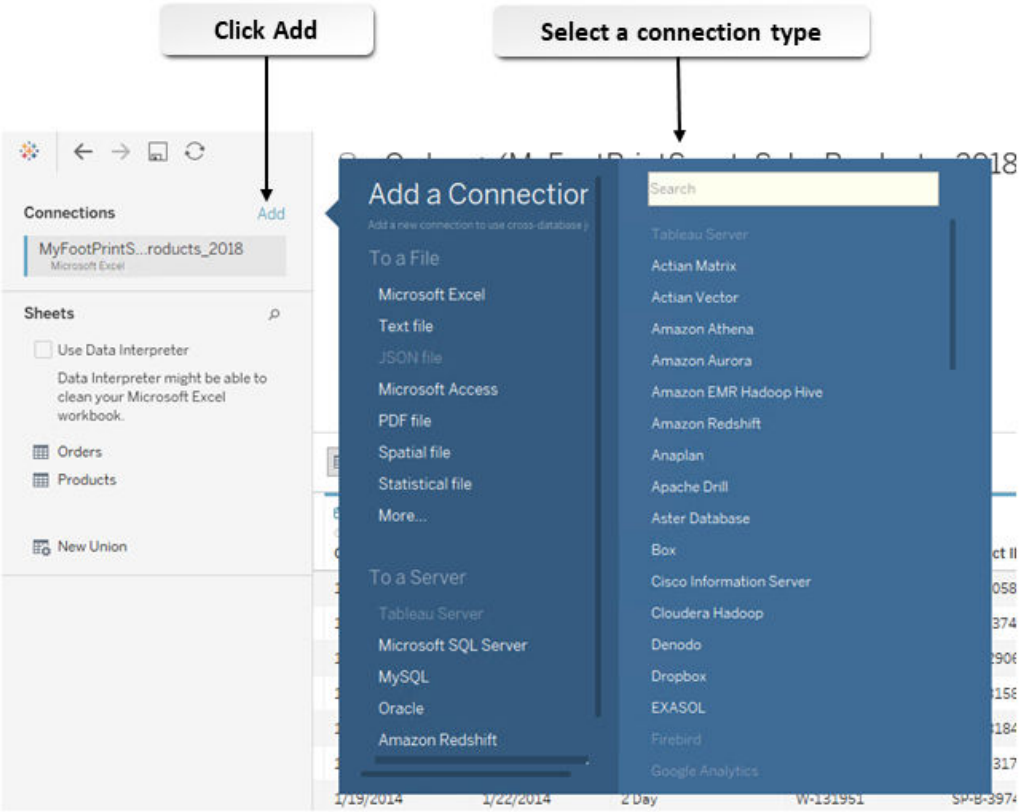



*Figure 2-7: Connections and cross-database joins.*

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Create Joins to Data Sources.** |

# ACTIVITY 2-1
## Creating Joins

### Data File
C:\095205Data\Joining Data\MyFootPrintSportsSalesProducts_2018.xlsx

### Before You Begin
Tableau Desktop is open.

### Scenario
The structure of your data source has changed and now your sales information has been split into three distinct parts. Data related to products has been split into a separate worksheet in Excel, and customer information is in a CSV file. You want to join these three data sources into one inside of Tableau. As a first step, you will join the sales and product worksheets in Excel and review which join type best fits your needs. You recently updated some of your product IDs and know the data source for products has some of those old IDs. You don't want to include those, so you won't use an outer join, and since the product data source will be the right table, you won't use a right join, either.

---

1. Review the Excel source files.
   a) Open File Explorer and navigate to **C:\095205Data\Joining Data**.
   b) Open **MyFootPrintSportsSalesProducts_2018.xlsx**.
   c) In Excel, on the **Orders** worksheet, review the data. This worksheet contains four years' worth of sales data. Each row is a line item of an order.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ID | Order ID | Order Date | Ship Date | Shipping Method | Customer ID | Product ID | Quantity | Sales | Profit | Shipping Cost |
| 2 | 11105 | 2013-38123 | 11/30/2013 | 12/3/2013 | Standard | R-108251 | AP-W-3236 | 2 | 27.98 | $8.39 | 3.20 |
| 3 | 11106 | 2013-38129 | 11/30/2013 | 12/4/2013 | Standard | S-118601 | AP-W-3248 | 1 | 13.99 | $4.20 | 2.60 |
| 4 | 12037 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | SP-T-3058 | 6 | $143.94 | $28.79 | 12.08 |
| 5 | 12038 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | AP-M-3748 | 6 | $71.94 | $28.78 | 7.04 |
| 6 | 12039 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | SP-K-2906 | 3 | $74.97 | $29.99 | 7.25 |
| 7 | 12040 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | SP-F-3158 | 2 | $57.98 | $17.39 | 6.06 |
| 8 | 12041 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | SP-F-3184 | 3 | $86.97 | $26.09 | 8.09 |
| 9 | 12045 | 2014-40923 | 1/15/2014 | 1/16/2014 | First Class | S-115901 | FW-H-3177 | 4 | $183.96 | $55.19 | 11.20 |
| 10 | 12051 | 2014-40927 | 1/19/2014 | 1/22/2014 | 2 Day | W-131951 | SP-B-3974 | 4 | $83.96 | $20.99 | 7.88 |
| 11 | 12076 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 | FW-H-3643 | 2 | $85.98 | $25.79 | 6.30 |
| 12 | 12077 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 | SP-K-6098 | 2 | $31.98 | $12.79 | 3.60 |
| 13 | 12098 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 | SP-S-3159 | 4 | $95.96 | $47.98 | 5.84 |
| 14 | 12099 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 | SP-K-2849 | 3 | $71.97 | $28.79 | 4.88 |
| 15 | 12105 | 2014-40947 | 2/8/2014 | 2/11/2014 | 2 Day | S-122501 | SP-T-2922 | 4 | $95.96 | $19.19 | 8.72 |

Orders | Products

d) In Excel, on the **Products** worksheet, review the data. This worksheet contains the different products sold by My Footprint Sports. Each row is a different product.

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | **Product ID** | **Category** | **SubCategory** | **Product Name** | **Unit Price** |
| 2 | AP-M-3100 | Apparel | Men | Men's Tank Top Winter Special | $10.99 |
| 3 | AP-M-3124 | Apparel | Men | Men's Tank Top Black | $11.99 |
| 4 | AP-M-3125 | Apparel | Men | Men's Tank Top White | $11.99 |
| 5 | AP-M-3200 | Apparel | Men | Men's Tank Top Summer Special | $10.99 |
| 6 | AP-M-3373 | Apparel | Men | Men's Tank Top Red | $11.99 |
| 7 | AP-M-3407 | Apparel | Men | Men's Tank Top Blue | $11.99 |
| 8 | AP-M-3748 | Apparel | Men | Men's Tank Top Green | $11.99 |
| 9 | AP-M-4140 | Apparel | Men | Men's Tank Top Yellow | $11.99 |
| 10 | AP-M-4142 | Apparel | Men | Men's Tank Top Orange | $11.99 |
| 11 | AP-M-4300 | Apparel | Men | Men's T-Shirt Black | $15.99 |
| 12 | AP-M-4788 | Apparel | Men | Men's T-Shirt White | $15.99 |
| 13 | AP-M-4799 | Apparel | Men | Men's T-Shirt Red | $15.99 |
| 14 | AP-M-4870 | Apparel | Men | Men's T-Shirt Blue | $15.99 |
| 15 | AP-M-4871 | Apparel | Men | Men's T-Shirt Green | $15.99 |

Orders | **Products**

e) Keep the Excel spreadsheet open.

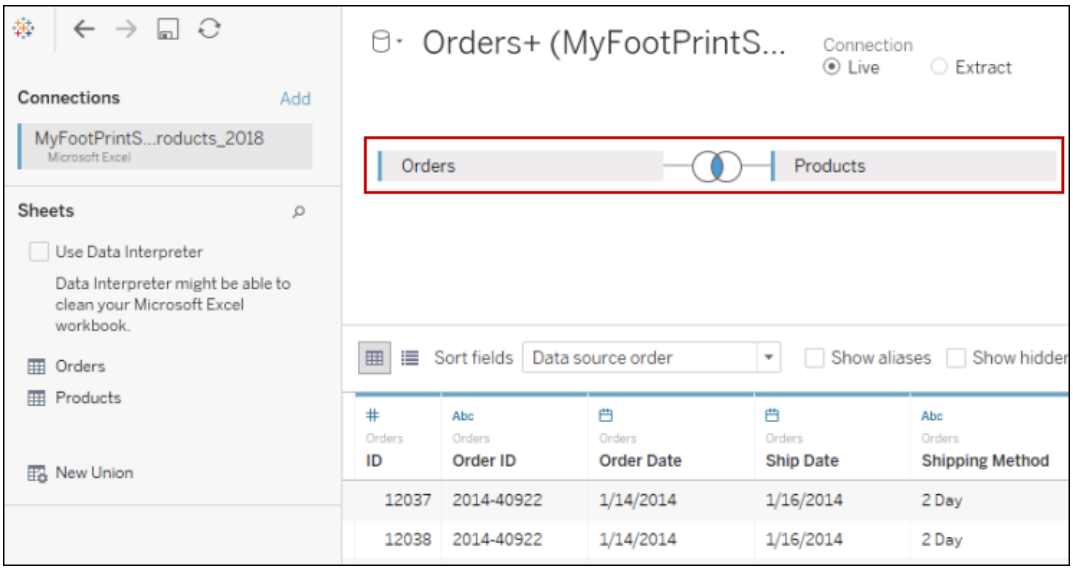
## 2. Connect to a data source.

a) In Tableau, in the **Connect** pane, under **To a File**, select **Microsoft Excel**.
b) In the **Open** dialog box, navigate to **C:\095205Data\Joining Data**.
c) Select **MyFootPrintSportsSalesProducts_2018.xlsx** and then select **Open**.

## 3. Create an inner join.

a) In the left pane, under **Sheets**, select and drag **Orders** to the canvas.
b) In the grid, review the data rows and columns.
c) In the left pane, under **Sheets**, select and drag **Products** to the canvas.

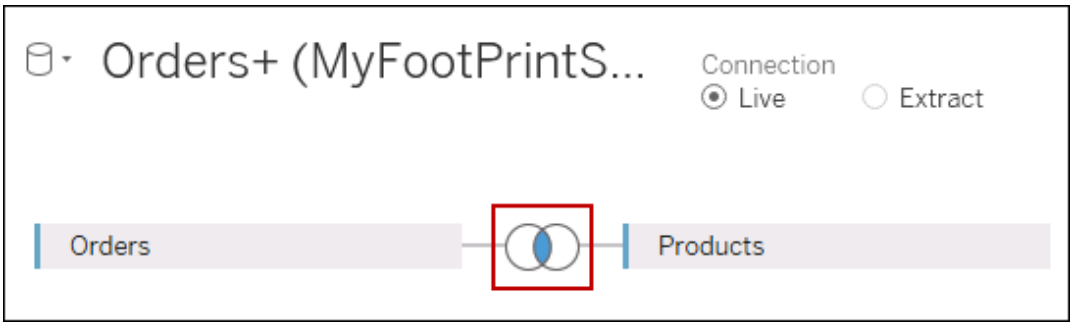
d) In the canvas, observe how Tableau automatically joined the two sheets.



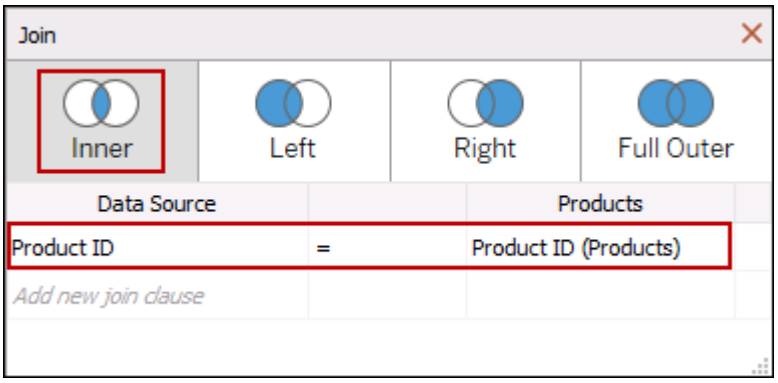

e) In the data grid, scroll to the right to view the fields added from the **Products** sheet.

4. Review the join type.

a) In the canvas, select the **Join** icon between the two sheets.



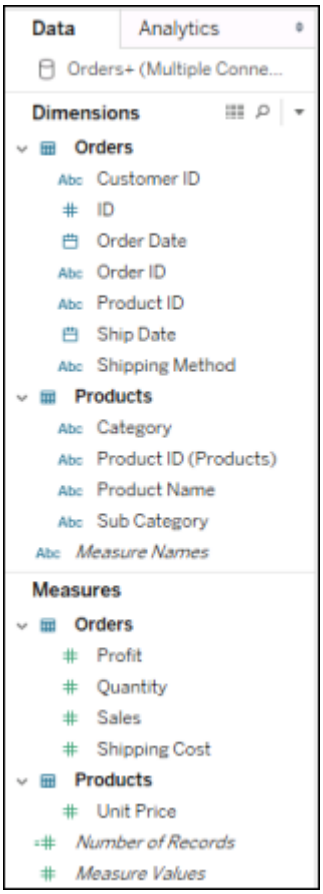

b) In the **Join** dialog box, observe that an inner join was implemented using the **Product ID** field in both sheets.




c) Close the **Join** dialog box.

d) At the bottom of the screen, select the **Sheet 1** worksheet.

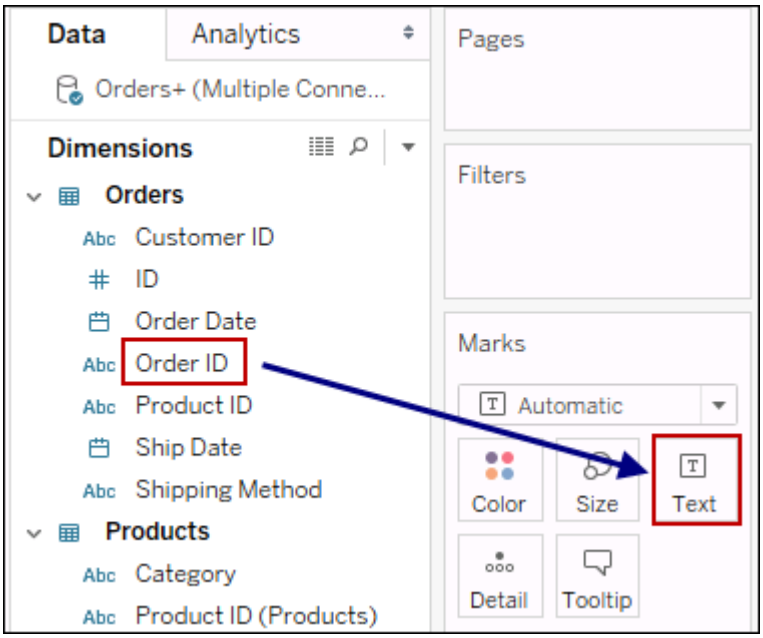
e) In the **Data** pane, observe that the combined fields from both sheets are listed.




5. Create an order view to review joins.

a) In the **Data** pane, under **Dimensions**, select and drag **Order ID** to the **Rows** shelf.

b) In the **Data** pane, under **Dimensions**, select and drag **Order ID** to the **Text** box in the **Marks** card.



c) In the **Marks** card, right-click **Order ID** and select **Measure→Count (Distinct)**.
d) From the menu, select **Analysis→Totals→Show Column Grand Totals**.
e) From the menu, select **Analysis→Totals→Column Totals to Top**.
f) In the canvas, observe the list of unique order IDs and the total number of orders listed at the top.

| Order ID | |
|---|---|
| Grand Total | 648 |
| 2014-40922 | 1 |
| 2014-40923 | 1 |
| 2014-40927 | 1 |
| 2014-40935 | 1 |
| 2014-40944 | 1 |

6. Rename the **Sheet 1** worksheet.
   a) Right-click the **Sheet 1** tab and select **Rename**.
   b) Type ***Orders*** and press **Enter**.

7. Create a product view to review joins.
   a) On the toolbar, select the **New Worksheet** button.
   b) In the **Data** pane, under **Dimensions**, under **Products**, select and drag **Product ID (Products)** to the **Rows** shelf.

   **Note:** Ensure you use the **Product ID (Products)** field under **Products**.

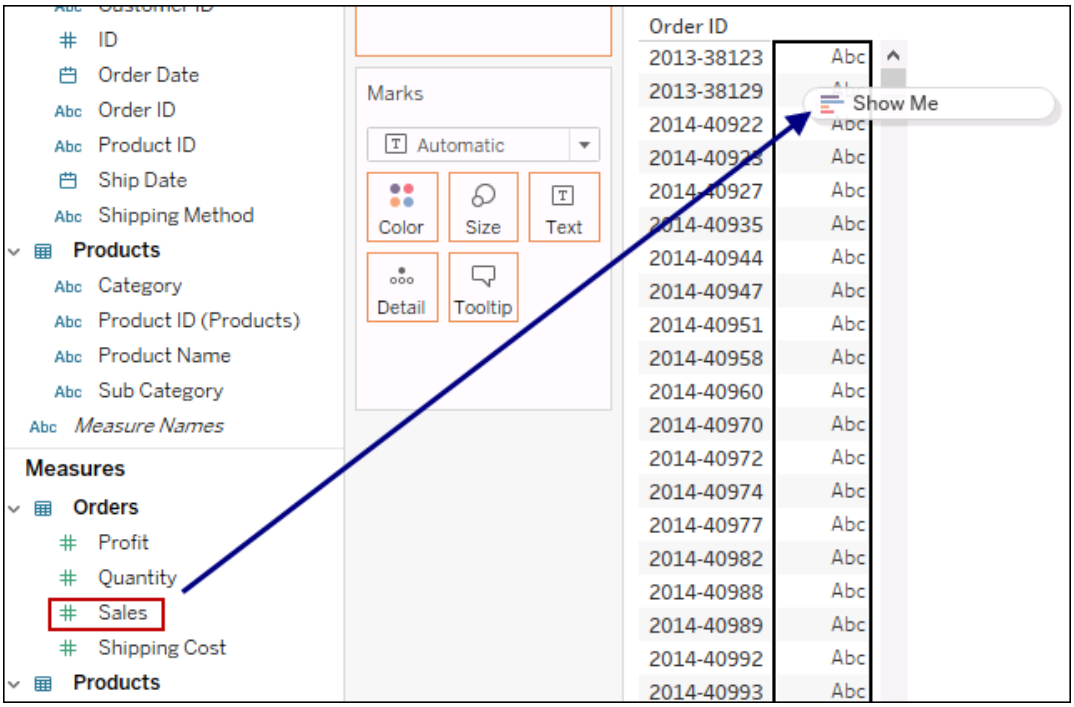
   c) In the **Data** pane, under **Dimensions**, under **Products**, select and drag **Product ID (Products)** to the **Text** box in the **Marks** card.
   d) In the **Marks** card, right-click **Product ID (Products)** and select **Measure→Count (Distinct)**.
   e) From the menu, select **Analysis→Totals→Show Column Grand Totals**.
   f) From the menu, select **Analysis→Totals→Column Totals to Top**.
   g) In the canvas, observe the list of unique product IDs and the total number of products listed at the top.

| Product ID (.. | |
|---|---|
| Grand Total | 264 |
| AP-M-3124 | 1 |
| AP-M-3125 | 1 |
| AP-M-3373 | 1 |
| AP-M-3407 | 1 |
| AP-M-3748 | 1 |

8. Rename the **Sheet 2** worksheet.
   a) Right-click the **Sheet 2** tab and select **Rename**.
   b) Type ***Products*** and press **Enter**.

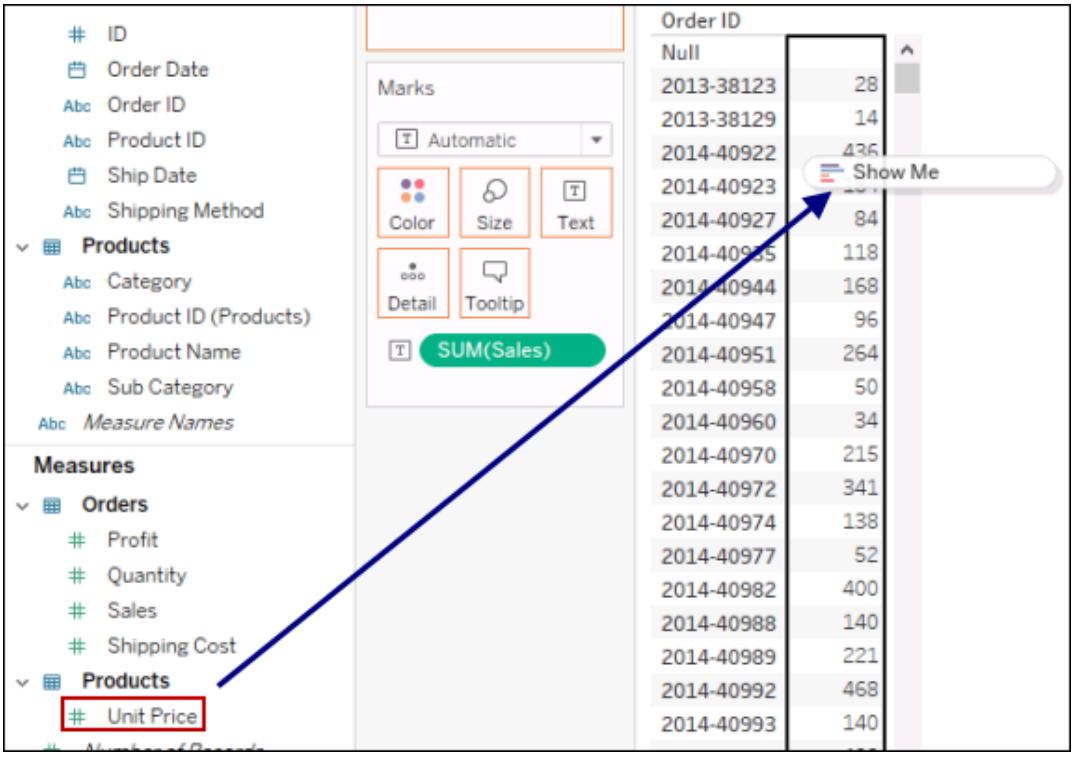
9. Create a combined view to review joins.
   a) On the toolbar, select the **New Worksheet** button.
   b) In the **Data** pane, under **Dimensions**, select and drag **Order ID** to the **Rows** shelf.

c) In the **Data** pane, under **Measures**, select and drag **Sales** to the canvas to the column to the right of **Order ID**.




| | |
|---|---|
| | **Note:** The **Show Me** button will appear when your mouse is over the correct position. |

d) In the **Data** pane, under **Measures**, select and drag **Unit Price** to the canvas to the **Sales** column.



e) In the canvas, observe the list of orders that has the sales amount from the **Orders** sheet and the **Unit Price** from the **Products** sheet.

## 10. Rename the **Sheet 3** worksheet.

a) Right-click the **Sheet 3** tab and select **Rename**.

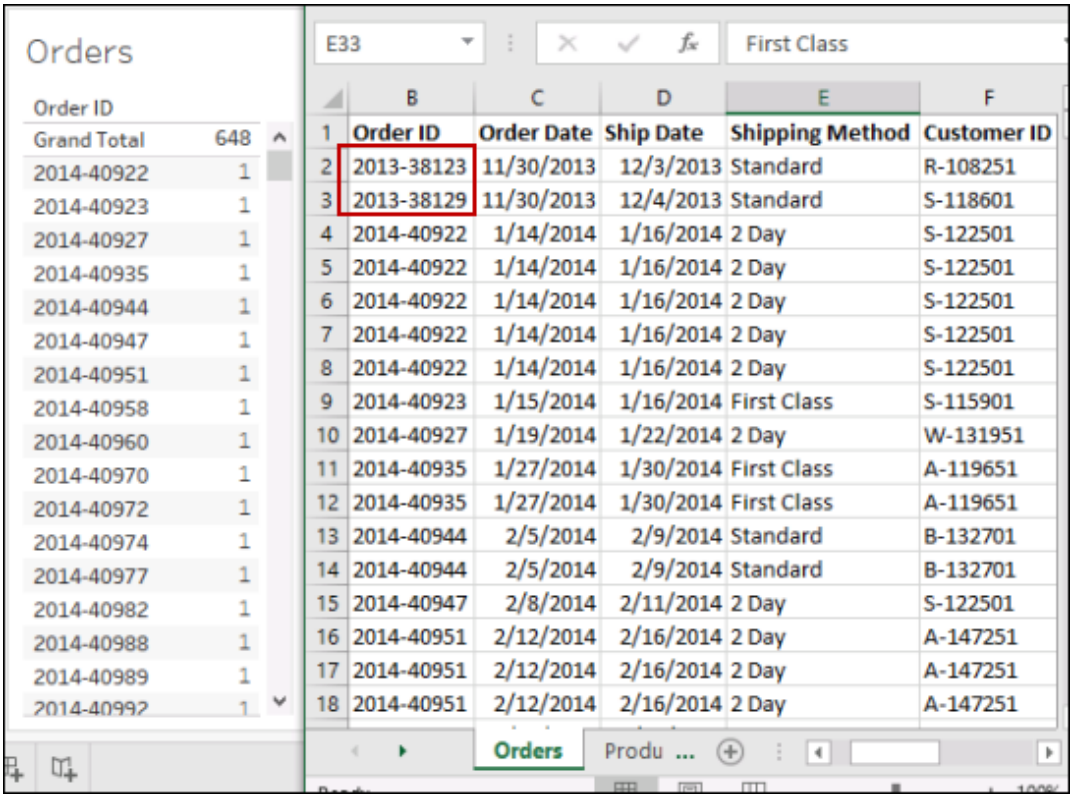
b) Type ***Joined Data*** and press **Enter**.

## 11. Compare inner join results.

a) At the bottom of the screen, select the **Orders** worksheet.

b) Restore the **MyFootPrintSportsSalesProducts_2018.xlsx** Excel spreadsheet and select the **Orders** sheet. Either arrange the Tableau and Excel windows so you can see the **Order ID** column in the spreadsheet and the **Order ID** column in the **Orders** view at the same time, or switch between the two windows to compare them.

c) Compare the **Orders** view to the **Orders** Excel worksheet and observe that the first two order IDs (**2013-38123** and **2013-38129**) from the first four records of the spreadsheet are not listed in the view. With an inner join, the **Product ID** must exist in both sheets; otherwise, the records related to that **Product ID** will not be included. These two orders contain products that are not in the **Products** sheet.

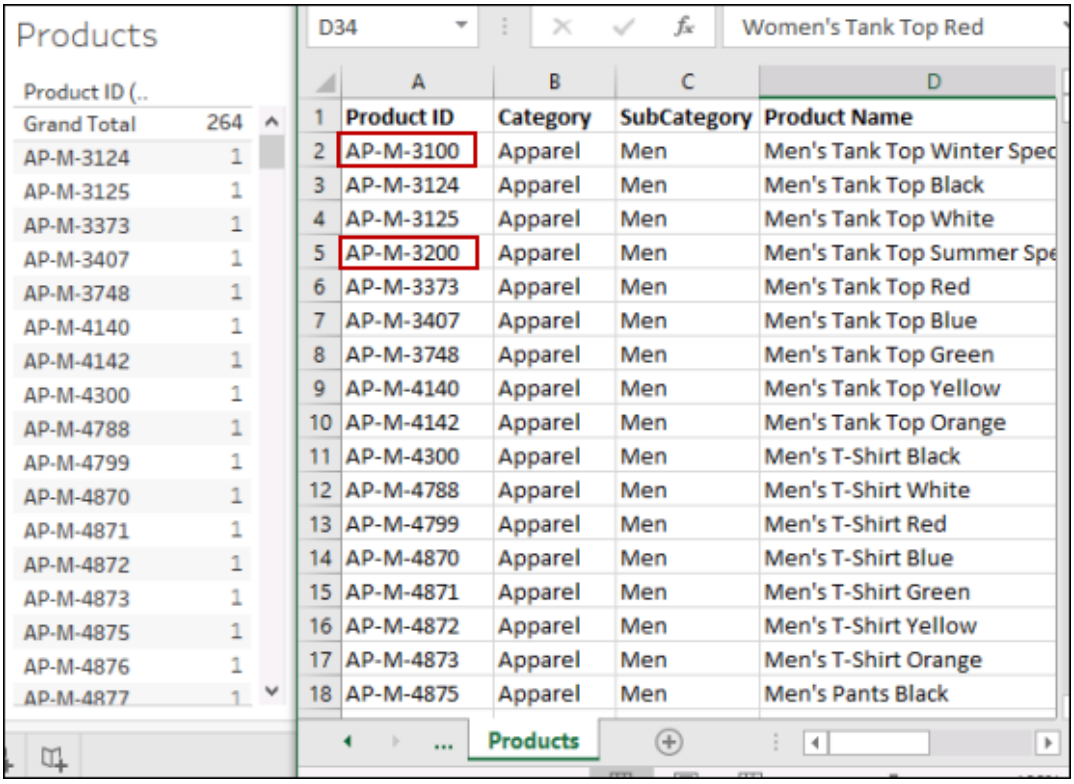



d) In Tableau, note the total number of order IDs in the view is **648**. This is the number of order line items you will have where the **Product ID** exists in both the **Orders** and **Products** tables. Selecting a join type that includes records from the **Orders** table that do not appear in the **Products** table will increase the number of order line items in Tableau.

e) At the bottom of the screen, select the **Products** worksheet.

f) Restore the **MyFootPrintSportsSalesProducts_2018.xlsx** Excel spreadsheet and select the **Products** sheet. Either arrange the Tableau and Excel windows so you can see the **Product ID** column in the spreadsheet and the **Product ID** column in the **Product** view at the same time, or switch between the two windows to compare them.

g) Compare the **Products** view to the **Products** Excel worksheet and observe that two product IDs (**AP-M-3100** and **AP-M-3200**) from the first four records of the spreadsheet are not listed in the view. These two products are not in the **Orders** sheet.



Products

| Product ID (.. | |
|---|---|
| Grand Total | 264 |
| AP-M-3124 | 1 |
| AP-M-3125 | 1 |
| AP-M-3373 | 1 |
| AP-M-3407 | 1 |
| AP-M-3748 | 1 |
| AP-M-4140 | 1 |
| AP-M-4142 | 1 |
| AP-M-4300 | 1 |
| AP-M-4788 | 1 |
| AP-M-4799 | 1 |
| AP-M-4870 | 1 |
| AP-M-4871 | 1 |
| AP-M-4872 | 1 |
| AP-M-4873 | 1 |
| AP-M-4875 | 1 |
| AP-M-4876 | 1 |
| AP-M-4877 | 1 |

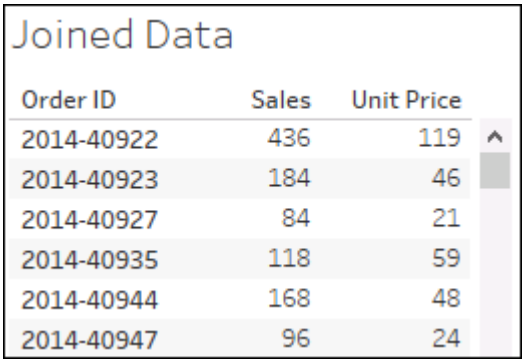
D34 | Women's Tank Top Red

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Product ID | Category | SubCategory | Product Name |
| 2 | AP-M-3100 | Apparel | Men | Men's Tank Top Winter Spec |
| 3 | AP-M-3124 | Apparel | Men | Men's Tank Top Black |
| 4 | AP-M-3125 | Apparel | Men | Men's Tank Top White |
| 5 | AP-M-3200 | Apparel | Men | Men's Tank Top Summer Spe |
| 6 | AP-M-3373 | Apparel | Men | Men's Tank Top Red |
| 7 | AP-M-3407 | Apparel | Men | Men's Tank Top Blue |
| 8 | AP-M-3748 | Apparel | Men | Men's Tank Top Green |
| 9 | AP-M-4140 | Apparel | Men | Men's Tank Top Yellow |
| 10 | AP-M-4142 | Apparel | Men | Men's Tank Top Orange |
| 11 | AP-M-4300 | Apparel | Men | Men's T-Shirt Black |
| 12 | AP-M-4788 | Apparel | Men | Men's T-Shirt White |
| 13 | AP-M-4799 | Apparel | Men | Men's T-Shirt Red |
| 14 | AP-M-4870 | Apparel | Men | Men's T-Shirt Blue |
| 15 | AP-M-4871 | Apparel | Men | Men's T-Shirt Green |
| 16 | AP-M-4872 | Apparel | Men | Men's T-Shirt Yellow |
| 17 | AP-M-4873 | Apparel | Men | Men's T-Shirt Orange |
| 18 | AP-M-4875 | Apparel | Men | Men's Pants Black |

h) In Tableau, note that the total number of product IDs in the view is **264**. This is the number of products you will have where the **Product ID** exists in both the **Orders** and **Products** tables. Selecting a join type that includes records from the **Products** table that do not appear in the **Orders** table will increase the number of products in Tableau.

i) At the bottom of the screen, select the **Joined Data** worksheet.

j) Observe that all of the orders listed have values for both **Sales** and **Unit Price**.



Joined Data

| Order ID | Sales | Unit Price |
|---|---|---|
| 2014-40922 | 436 | 119 |
| 2014-40923 | 184 | 46 |
| 2014-40927 | 84 | 21 |
| 2014-40935 | 118 | 59 |
| 2014-40944 | 168 | 48 |
| 2014-40947 | 96 | 24 |

## 12. Change to a left join.

a) At the bottom of the screen, select the **Data Source** tab.

b) In the canvas, select the **Join** icon between the two sheets.

c) In the **Join** dialog box, select **Left**.

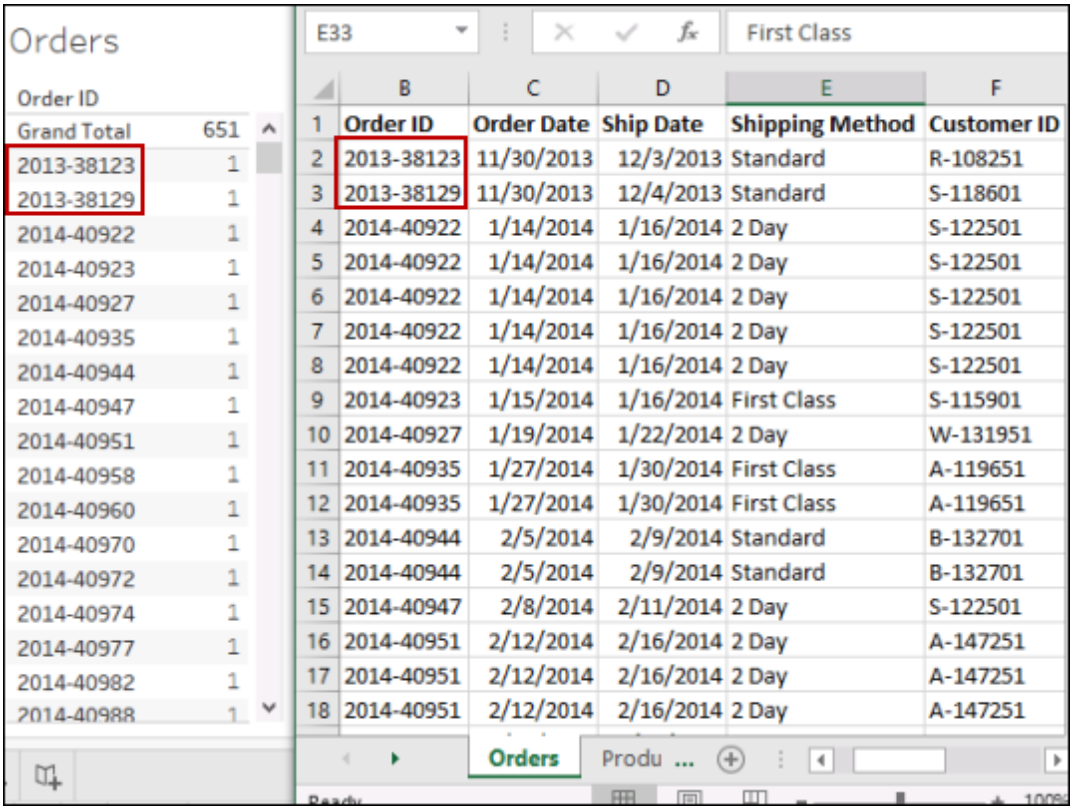
d) Close the **Join** dialog box.

e) At the bottom of the screen, select the **Orders** worksheet.

13. Compare left join results.

a) Restore the **MyFootPrintSportsSalesProducts_2018.xlsx** Excel spreadsheet and select the **Orders** sheet.

b) Compare the **Orders** view to the **Orders** Excel worksheet and observe that it appears that all of the order IDs from the **Orders** sheet are listed in the view. With a left join, all of the records from the first sheet are included, but only records from the second sheet that match a **Product ID** in the first sheet are included. So in this case, all of the records from the **Orders** sheet are included and only those in the **Products** sheet are included if there is a match in the **Orders** sheet.



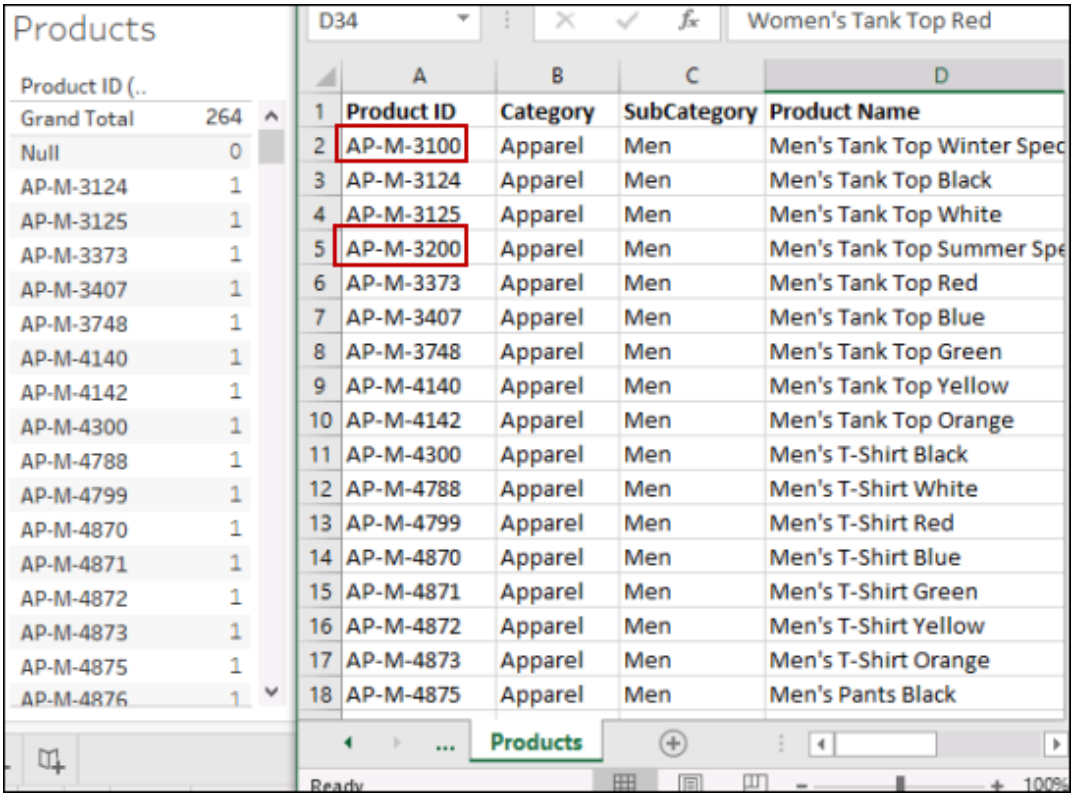

| | B | C | D | E | F |
|---|---|---|---|---|---|
| 1 | Order ID | Order Date | Ship Date | Shipping Method | Customer ID |
| 2 | 2013-38123 | 11/30/2013 | 12/3/2013 | Standard | R-108251 |
| 3 | 2013-38129 | 11/30/2013 | 12/4/2013 | Standard | S-118601 |
| 4 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 |
| 5 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 |
| 6 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 |
| 7 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 |
| 8 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 |
| 9 | 2014-40923 | 1/15/2014 | 1/16/2014 | First Class | S-115901 |
| 10 | 2014-40927 | 1/19/2014 | 1/22/2014 | 2 Day | W-131951 |
| 11 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 |
| 12 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 |
| 13 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 |
| 14 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 |
| 15 | 2014-40947 | 2/8/2014 | 2/11/2014 | 2 Day | S-122501 |
| 16 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 |
| 17 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 |
| 18 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 |

c) In Tableau, note the total number of order IDs in the view is **651**, three more than with the inner join.

d) At the bottom of the screen, select the **Products** worksheet.

e) Restore the **MyFootPrintSportsSalesProducts_2018.xlsx** Excel spreadsheet and select the **Products** sheet.

f) Compare the **Products** view to the **Products** Excel worksheet and observe that two product IDs (**AP-M-3100** and **AP-M-3200**) from the first four records of the spreadsheet are not listed in the view. These two products are not in the **Orders** sheet.



Products

| Product ID (.. | |
|---|---|
| Grand Total | 264 |
| Null | 0 |
| AP-M-3124 | 1 |
| AP-M-3125 | 1 |
| AP-M-3373 | 1 |
| AP-M-3407 | 1 |
| AP-M-3748 | 1 |
| AP-M-4140 | 1 |
| AP-M-4142 | 1 |
| AP-M-4300 | 1 |
| AP-M-4788 | 1 |
| AP-M-4799 | 1 |
| AP-M-4870 | 1 |
| AP-M-4871 | 1 |
| AP-M-4872 | 1 |
| AP-M-4873 | 1 |
| AP-M-4875 | 1 |
| AP-M-4876 | 1 |

D34 | Women's Tank Top Red

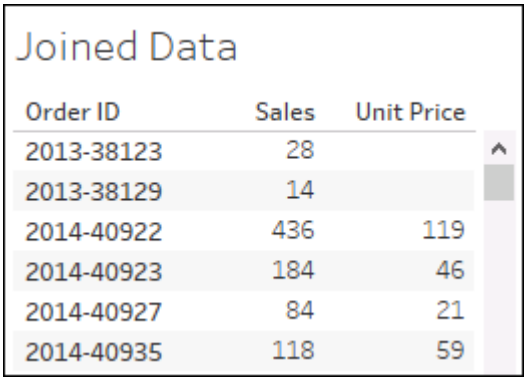
| | A | B | C | D |
|---|---|---|---|---|
| 1 | Product ID | Category | SubCategory | Product Name |
| 2 | AP-M-3100 | Apparel | Men | Men's Tank Top Winter Spec |
| 3 | AP-M-3124 | Apparel | Men | Men's Tank Top Black |
| 4 | AP-M-3125 | Apparel | Men | Men's Tank Top White |
| 5 | AP-M-3200 | Apparel | Men | Men's Tank Top Summer Spe |
| 6 | AP-M-3373 | Apparel | Men | Men's Tank Top Red |
| 7 | AP-M-3407 | Apparel | Men | Men's Tank Top Blue |
| 8 | AP-M-3748 | Apparel | Men | Men's Tank Top Green |
| 9 | AP-M-4140 | Apparel | Men | Men's Tank Top Yellow |
| 10 | AP-M-4142 | Apparel | Men | Men's Tank Top Orange |
| 11 | AP-M-4300 | Apparel | Men | Men's T-Shirt Black |
| 12 | AP-M-4788 | Apparel | Men | Men's T-Shirt White |
| 13 | AP-M-4799 | Apparel | Men | Men's T-Shirt Red |
| 14 | AP-M-4870 | Apparel | Men | Men's T-Shirt Blue |
| 15 | AP-M-4871 | Apparel | Men | Men's T-Shirt Green |
| 16 | AP-M-4872 | Apparel | Men | Men's T-Shirt Yellow |
| 17 | AP-M-4873 | Apparel | Men | Men's T-Shirt Orange |
| 18 | AP-M-4875 | Apparel | Men | Men's Pants Black |

Products | Ready | 100%

g) In Tableau, note the total number of product IDs in the view is still **264** and the null value. The null value represents a **Product ID** in the **Orders** table that does not exist in the **Products** table.

h) At the bottom of the screen, select the **Joined Data** worksheet.

i) Observe that the first two orders listed do not have values for **Unit Price** because there isn't a record for them in the **Products** sheet.



Joined Data

| Order ID | Sales | Unit Price |
|---|---|---|
| 2013-38123 | 28 | |
| 2013-38129 | 14 | |
| 2014-40922 | 436 | 119 |
| 2014-40923 | 184 | 46 |
| 2014-40927 | 84 | 21 |
| 2014-40935 | 118 | 59 |

## 14. Select an inner join.

a) At the bottom of the screen, select the **Data Source** tab.

b) In the canvas, select the **Join** icon between the two sheets.

c) In the **Join** dialog box, select **Inner**.
To avoid having to resolve incomplete product records, you will use an inner join to only keep records that have matches.

d) Close the **Join** dialog box.

e) Close Excel.

15. Save the workbook.

a) From the menu, select **File→Save As**.
b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Joining Data** folder.
c) In the **File name** box, type *MyCopy_MyFootPrintSports_2018_2* and select **Save**.

# TOPIC B

## Troubleshoot Joins

Joining data can introduce issues with the combined data. In this topic, you will troubleshoot joins to solve any problems with data and ensure combined data is accurate.

### Data Grid

The ***data grid*** on the **Data Sources** page is used to configure the information coming in from data sources and to adjust the ***metadata*** properties about the data being brought into Tableau.

When you create joins and cross-database joins, the data grid is the first place to look to see if data has been combined in the way you expected and meets your data analysis requirements. The first 1,000 rows of data is displayed in the data grid, providing you with a large sample to work with. Here, you can sort data, hide columns, rename fields, and change field types. You can also use the data grid to locate unexpected columns and null values in the data. You can create calculations to attempt to address fields containing nulls, or to create new fields with data from the joined data sources. You can also perform general data preparation tasks, such as cleaning data with **Data Interpreter**.

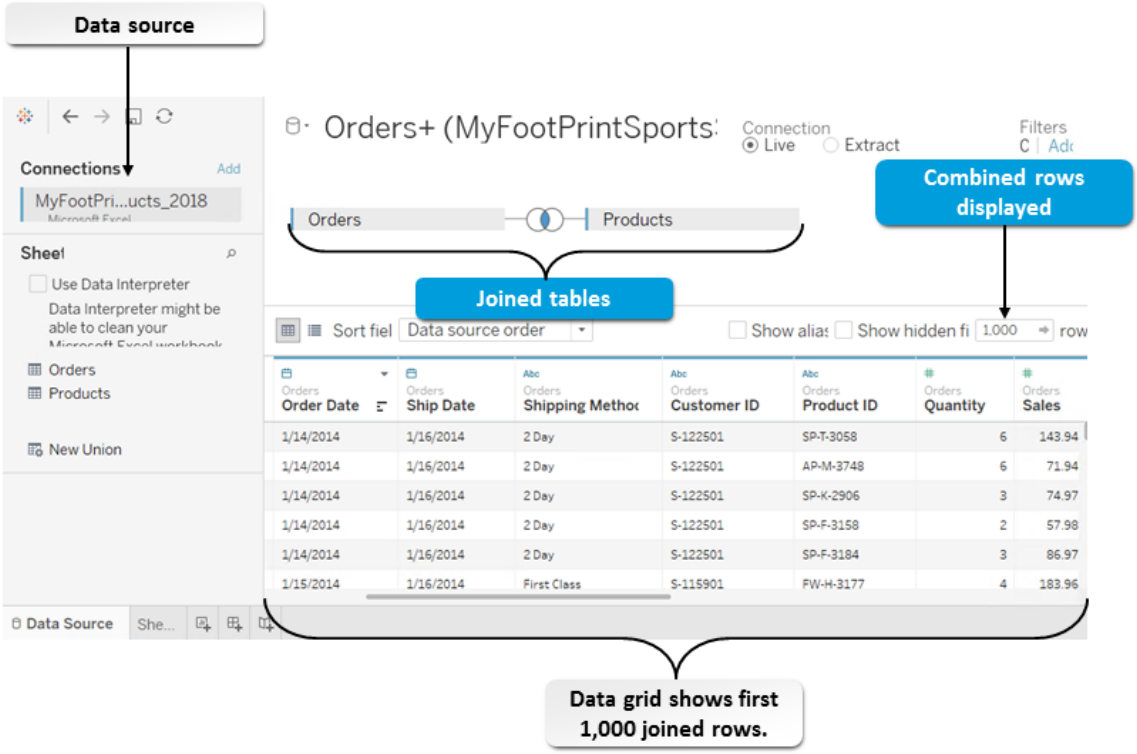



*Figure 2-8: The data grid.*

#### Data Grid vs. Metadata Grid

Metadata is information that describes data. For example, some common types of metadata used when working with data sources are field names and field types, such as numeric, string, and data. The data grid is used to configure relational data sources, while the ***metadata grid*** is used to configure multidimensional data sources (cubes) and extracts, and to adjust field properties for any type of connection.

## Troubleshoot Join Results

If the results you see in the data grid after a join have issues, you will need to take steps to troubleshoot the issues with the join. The following table explains some of the issues that might present themselves and the potential solutions.

| Issue | Description | Potential Corrective Action |
|---|---|---|
| No data in the data grid. | Tableau could not find matching rows to combine using the joined field. | Select a different join field or join condition. |
| Duplicate data. | Joined databases may contain data at different levels of detail. Since Tableau's default method of aggregation is to SUM, all numbers will be rolled up. | Consider changing the aggregation used, creating a calculation, or use data blending instead of joins. |
| Missing data. | Data columns you expect to see aren't present. This may indicate mismatched field names or types. | Review data sources and look for mismatches. Consider selecting a different field to join on, and a different join type. Use calculated fields to correct mismatches. |
| Many null values. | Many fields have no data to combine and therefore contain null values. | Consider switching to another join type. Specifically, if using an outer join, consider switching to an inner join. |
| One table has all null values. | There are no matches for the joined field for the table with the nulls. | Change the join field or join type. |

*Figure 2-9: Troubleshoot join results.*

## Fix Joined Data Issues with Join Calculations

When there is a mismatch between the data type in the joined fields, such as one field that is a date field and another field that is a string, the data grid may not show any data—or very little data—after the join is performed. In fact, one of the most common reasons for mismatched fields is formatting differences in date and string fields. In many cases like this, you can create a join calculation to fix the problem.

Join calculations can modify the format of the field in the tables you are joining. All functions except aggregation and table calculation functions are available for join calculations. You can create a join relationship by clicking the broken join and then clicking the field you wish to modify and selecting **Create Join Calculation.**

Some of the common scenarios where calculations are used to fix broken joins are as follows:

- **String mismatch.** When one table has data in a single field, such as Name, while the joining table has data in two fields, such as First Name and Last Name. You can use a calculation to merge the First Name and Last Name fields together into a Name field, or you can split the First Name and Last Name out into separate fields by separating the contents of the field at the space.
- **Date mismatch.** In this case, dates in separate tables are captured at different levels of detail. For example, one table may store dates by Month-Day-Year, while another may store dates as Month-Year. You can use date functions on both sides of the join to ensure rows are combined at the closest match.

## Troubleshoot Broken Joins

You should make sure the fields used to create a cross-database join have matching data types. If the field types don't match, you will see a red exclamation point indicating a broken join. To fix a broken join, try the following:

- Change the format of a key field in the join by using a calculation.
- Change the format of one of the fields in a text or spreadsheet data source from the data grid by using a data type menu.
- Change the format by using a convert to custom SQL option for one of the joined fields.

## Guidelines for Using Data Blending vs. Joins

**Note:** All of the Guidelines for this lesson are available as checklists from the **Checklist** tile on the CHOICE Course screen.

### Using Data Blending and Joins

When combining data in Tableau, consider these guidelines for when to use data blending and joins:

- Each scenario for combining data is different. Consider all of the following when deciding to join or blend data:
  - Data sources in use, as not all sources support joins or blends.
  - The number of connections in the workbook.
  - The number of records in the data.
- Consider experimenting with both joins and blends to see which approach yields the best combined data.
- Consider joining data:
  - **If adding data from an existing data source or connection.** Joins were designed for linking tables in the same data source and to bring new data in from the same data source.
  - **To replace multiple connections to the same data source.** Using a join rather than multiple connections to the same data source can improve performance and provide additional capabilities for filtering data.
  - **If you wish to publish the connection as a data source on Tableau Server or Tableau Online.** Joined data sources can be published, while blends cannot.
- Consider blending data:
  - **When data sources don't support cross-database joins.** Cubes and some extract-based connections—such as Salesforce—don't support cross-database joins. Blending is your only option in these cases.
  - **If the data you're combining is at different levels of detail.** When blending data in Tableau, data from secondary sources is aggregated at the level of the fields being joined, which in many cases, addresses issues with differing levels of detail.
  - **If combining data from more than one source.** Data blending is designed to combine data from different sources, whereas joins have traditionally been used to join data in the same database.
  - **If data is not clean after a join**. If, after a join, data does not match and results in extra columns and lots of null values, consider setting up both tables as separate data sources. You can then adjust the tables in the data grid by sorting, renaming, changing data types, and creating calculations. Once you've gotten the tables to substantially match, you can blend the data.
  - **If a join creates a lot of duplicated data.** This can be an indicator of data in joined tables being at different levels of detail. You can use calculations or changes in aggregation to make a join work, but blending the data may be an easier approach.
  - **If joining large data sets.** Joining large tables can affect the performance of the underlying database. If this is the case, consider blending the data. Since blending combines data, it often has been aggregated. There is less data to combine and it reduces the performance impact.

**Note:** How large of a data set is too large? The answer depends on the speed of the underlying hardware at the database server and Tableau Desktop. There are a number of other optimizations that may or may not be in place. In the end, you'll have to run some tests, such as running queries and letting Tableau compute the layout for a visualization. You will have to determine if those tasks are performing acceptably.

- **If you wish to show summary data and detail data in the same view.**

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Create Cross-Database Joins.**

# ACTIVITY 2-2
## Creating Cross-Database Joins

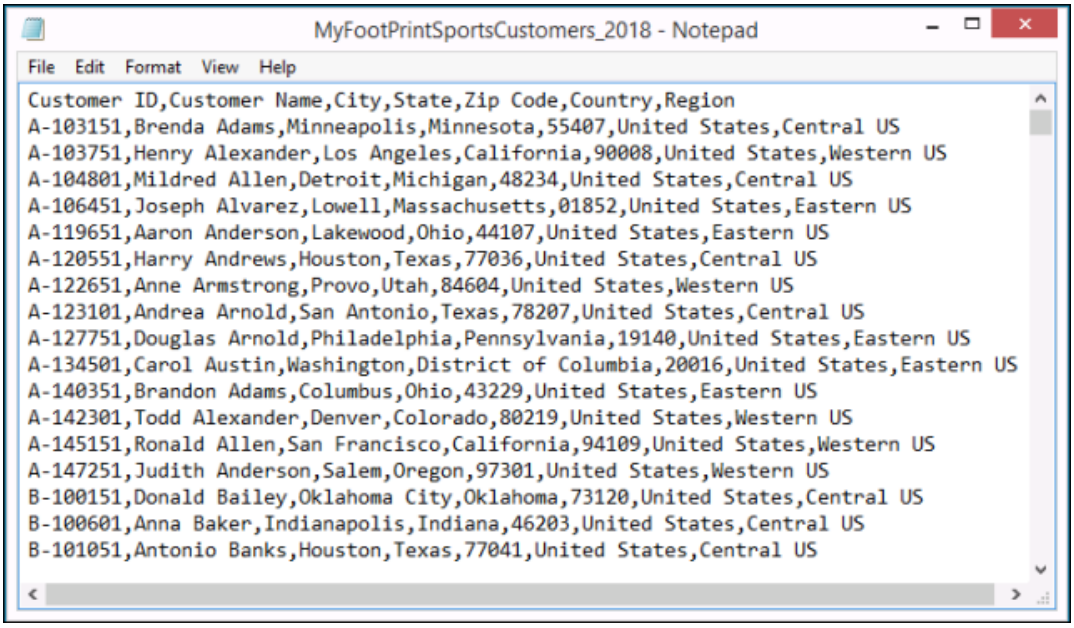
### Data File
C:\095205Data\Joining Data\MyFootPrintSportsCustomers_2018.csv

### Before You Begin
Tableau Desktop is open.

### Scenario
Now that you have joined your order and product data sources together, you want to join the customer data source. For now, you only want to work with a complete set of data, so you will use an inner join with the customer data and products.

---

1. Review the source files.
   a) In File Explorer, in the **C:\095205Data\Joining Data** folder, open **MyFootPrintSportsCustomers_2018.csv** with Notepad.
   b) In Notepad, review the data. This CSV file contains data on customers.

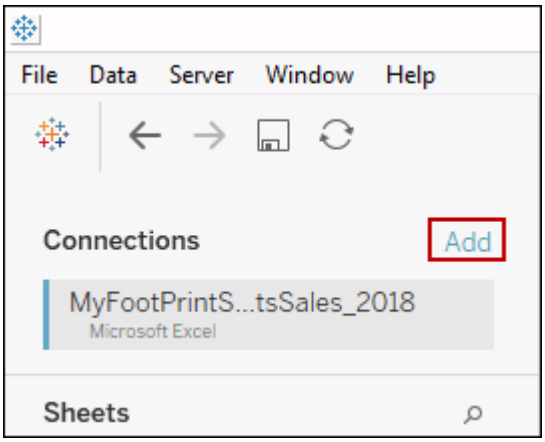



```
Customer ID,Customer Name,City,State,Zip Code,Country,Region
A-103151,Brenda Adams,Minneapolis,Minnesota,55407,United States,Central US
A-103751,Henry Alexander,Los Angeles,California,90008,United States,Western US
A-104801,Mildred Allen,Detroit,Michigan,48234,United States,Central US
A-106451,Joseph Alvarez,Lowell,Massachusetts,01852,United States,Eastern US
A-119651,Aaron Anderson,Lakewood,Ohio,44107,United States,Eastern US
A-120551,Harry Andrews,Houston,Texas,77036,United States,Central US
A-122651,Anne Armstrong,Provo,Utah,84604,United States,Western US
A-123101,Andrea Arnold,San Antonio,Texas,78207,United States,Central US
A-127751,Douglas Arnold,Philadelphia,Pennsylvania,19140,United States,Eastern US
A-134501,Carol Austin,Washington,District of Columbia,20016,United States,Eastern US
A-140351,Brandon Adams,Columbus,Ohio,43229,United States,Eastern US
A-142301,Todd Alexander,Denver,Colorado,80219,United States,Western US
A-145151,Ronald Allen,San Francisco,California,94109,United States,Western US
A-147251,Judith Anderson,Salem,Oregon,97301,United States,Western US
B-100151,Donald Bailey,Oklahoma City,Oklahoma,73120,United States,Central US
B-100601,Anna Baker,Indianapolis,Indiana,46203,United States,Central US
B-101051,Antonio Banks,Houston,Texas,77041,United States,Central US
```

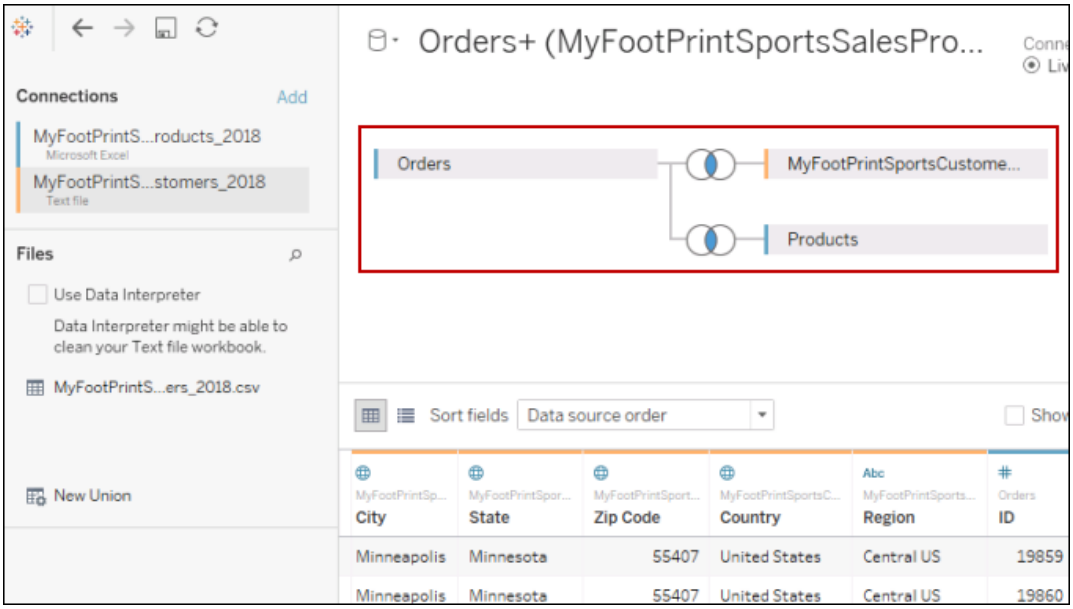
   c) Close Notepad.
   d) If prompted to save changes, select **Don't Save**.

2. Create a cross-database join.
   a) In Tableau, at the bottom of the screen, verify that the **Data Source** tab is selected.

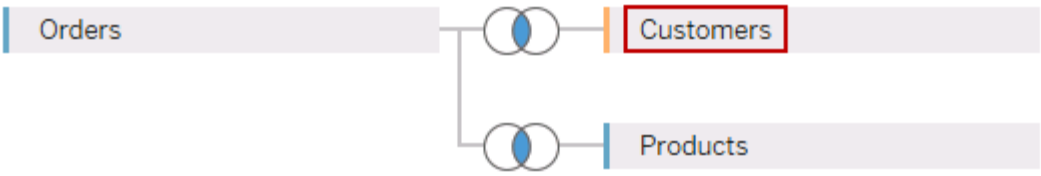
b) In the **Connections** pane, select **Add**.




c) In the **Add a Connection** context menu, under **To a File**, select **Text file**.
d) In the **Open** dialog box, select **MyFootPrintSportsCustomers_2018.csv** and then select **Open**.
e) In the canvas, observe how Tableau automatically joined the **Customers** data source.



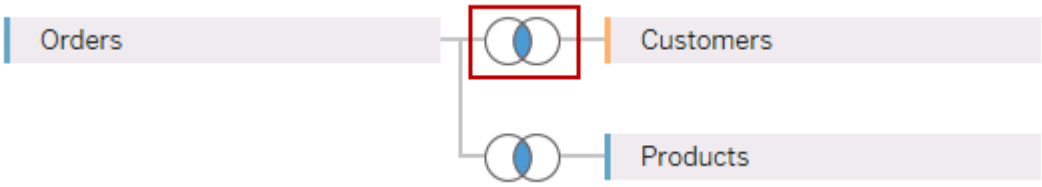

f) In the canvas, double-click **MyFootPrintSportsCustomers_2018** to select the text.
g) Type ***Customers*** and press **Enter** to rename the top label.




h) In the data grid, scroll to the right and back to the left to view the fields added from the **Customers** data source.

3. Review the join type.

a) In the canvas, select the **Join** icon between **Orders** and **Customers**.




b) In the **Join** dialog box, observe that an inner join was implemented using the **Customer ID** field in both data sources.
c) Close the **Join** dialog box.

## 4. Add customer data to a view.

a) At the bottom of the screen, select the **Joined Data** worksheet.
b) In the **Data** pane, under **Dimensions**, under **Customers**, select and drag **Region** to the **Rows** shelf.
c) In the canvas, observe the addition of **Region** to the table.

Joined Data

| Order ID | Region | Sales | Unit Price |
|---|---|---|---|
| 2014-40922 | Central US | 436 | 119 |
| 2014-40923 | Southern US | 184 | 46 |
| 2014-40927 | Western US | 84 | 21 |
| 2014-40935 | Eastern US | 118 | 59 |
| 2014-40944 | Central US | 168 | 48 |
| 2014-40947 | Central US | 96 | 24 |
| 2014-40951 | Western US | 264 | 111 |

# ACTIVITY 2-3

## Creating Views with Cross-Database Data

### Before You Begin

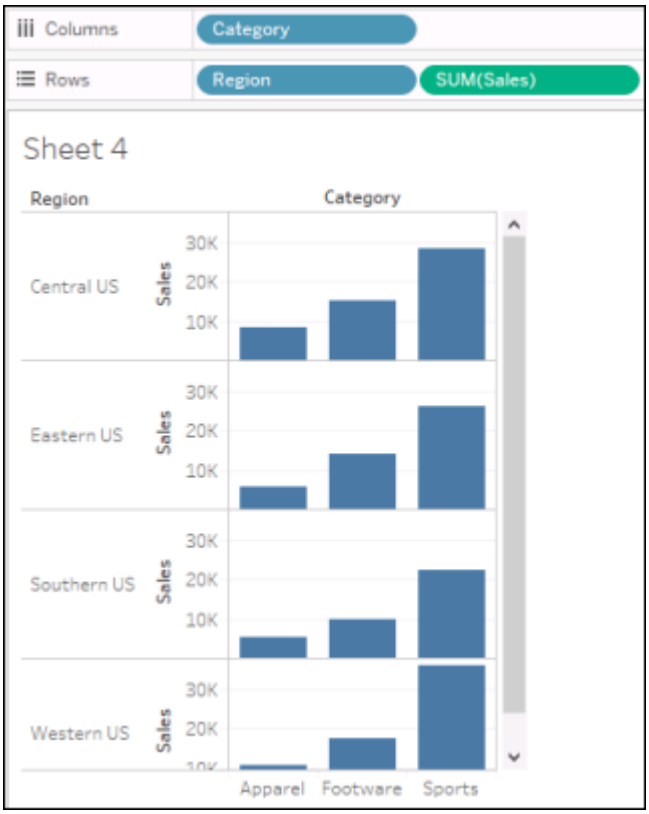
Tableau Desktop is open.

### Scenario

Now that you have joined your three data sources in Tableau, you will create views using that data. You want a view that shows sales for each region by category. You also want a view that shows profit for each subcategory by quarter for customers that had $1,000 or more in sales.

---

1. Create a view that shows sales for each region by category.
   a) On the toolbar, select the **New Worksheet** button.
   b) In the **Data** pane, under **Dimensions→Customers**, select and drag **Region** to the **Rows** shelf.
   c) In the **Data** pane, under **Dimensions→Products**, select and drag **Category** to the **Columns** shelf.
   d) In the **Data** pane, under **Measures→Orders**, select and drag **Sales** to the **Rows** shelf.
   e) In the canvas, observe the view that shows **Region** from the **Customers** source, **Category** from the **Products** source, and **Sales** from the **Orders** source.

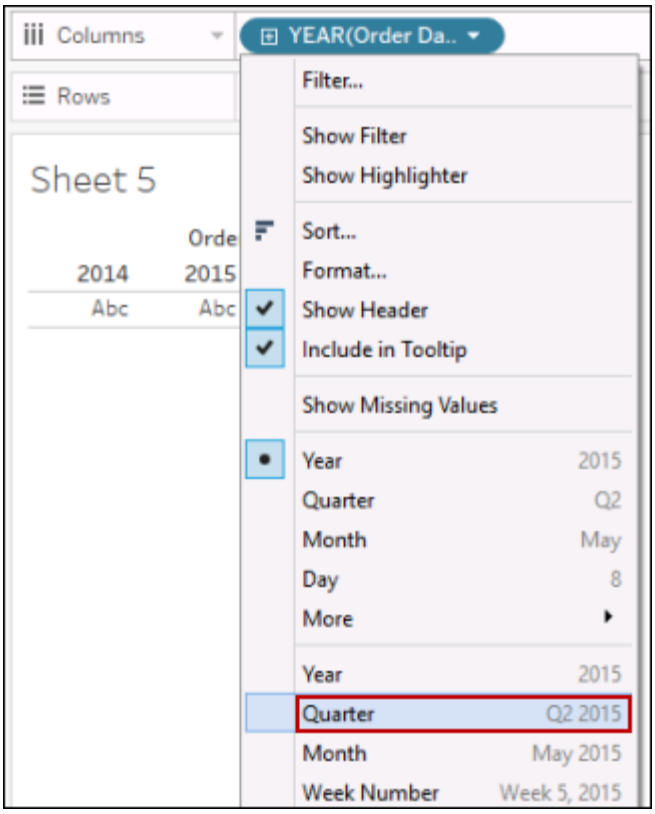



2. Rename the **Sheet 4** worksheet.
   a) Right-click the **Sheet 4** tab and select **Rename**.

b) Type ***Sales by Region and Category*** and press **Enter**.

## 3. Create a view that shows profit for each subcategory by quarter.

a) On the toolbar, select the **New Worksheet** button.
b) In the **Data** pane, under **Dimensions→Orders**, select and drag **Order Date** to the **Columns** shelf.
c) In the **Columns** shelf, right-click **YEAR(Order Date)** and select the second instance of **Quarter**.

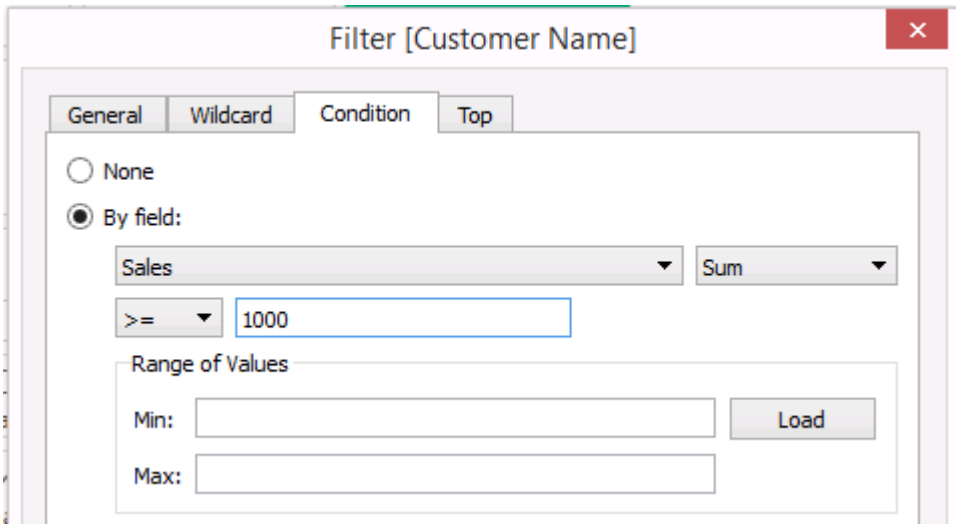



d) In the **Data** pane, under **Measures→Orders**, select and drag **Profit** to the **Rows** shelf.
e) In the **Data** pane, under **Dimensions→Products**, select and drag **Sub Category** to the **Color** box in the **Marks** card.

## 4. Filter the view to include customers that had $1,000 or more in sales only.

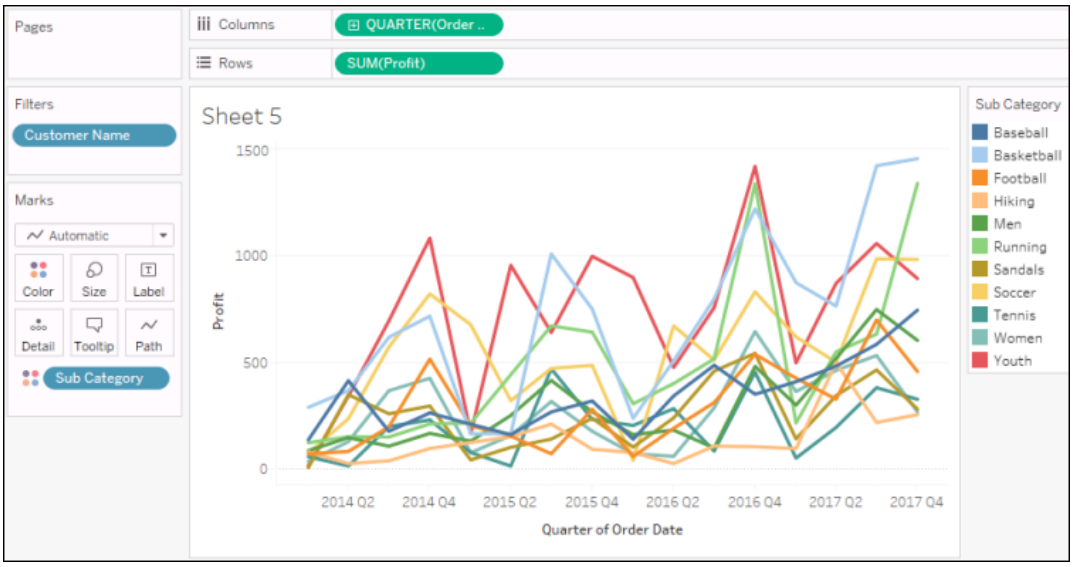
a) In the **Data** pane, under **Dimensions→Customers**, select and drag **Customer Name** to the **Filters** card.
b) In the **Filter [Customer Name]** dialog box, select the **Condition** tab.
c) Select **By field**.
d) In the **Profit** drop-down list, select **Sales**.
e) In the **=** drop-down list, select **>=**.

f) In the box, type *1000*




g) Select **OK**.

h) In the canvas, observe the view that shows profit for each subcategory over four years for customers that had $1,000 or more in sales during that time period.




5. Rename the **Sheet 5** worksheet.

a) Right-click the **Sheet 5** tab and select **Rename**.

b) Type ***Profit by Subcategory*** and press **Enter**.

# TOPIC C

## Merge Data with Unions

There are times when data analysts need to combine like data sets to perform comprehensive analysis. In this topic, you will use unions to merge data to visualize it in Tableau.

### Unions

Joins link tables at the row level, combining data by adding columns using data from the joined tables. This increases the amount of information per row of data. Performing a union adds new rows to an existing table, and can be thought of as something similar to a merge or update of a table. Use a union when you have similar data stored in multiple different tables, files, locations. or databases.

For example, each store in a nationwide company may keep its own sales records and you may need to unite tables from each store in the sales database to get a view of sales for all stores.

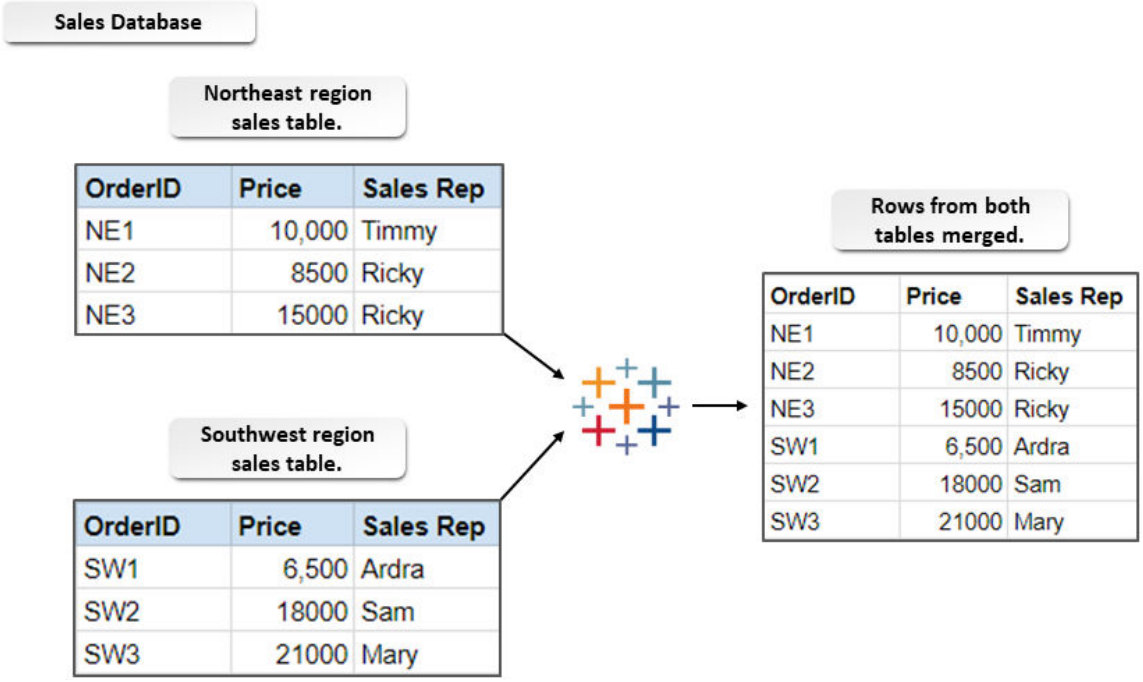



*Figure 2-10: A union.*

You can perform a union between two or more tables that are part of the same data source connection. Not all data sources support performing a union. If a union is supported, on the **Data Source** page, in the left pane, you can select **New Union.**

Before performing a union, you should verify that the tables involved in the union have the same number of fields. You should also verify that field names and field data types are the same. This may necessitate some data preparation before performing the union. If you have mismatched names or data types, you will have to perform some additional steps during the union process to clean up or merge fields.

Tables are combined by matching the names of fields when working with databases, or the names of column headers when working with spreadsheets, text files, PDF, or JSON files. If no matching field names or column headers exist, you can have Tableau perform a union based on the order of

the fields. Do this from the **Union** drop-down menu by selecting the option to **Generate field names automatically**.

When creating a union, you can drag-and-drop tables to add and remove them from the union. You can also see information about mismatched fields. Once a union is created, metadata is displayed showing you information about the tables in the union, the resulting fields, and any mismatched fields. A table name column is also added that shows the table name or sheet from which the data originated. You can make adjustments for mismatched fields and make other changes before updating your table with changes.

### Data Sources That Support Unions

The following data sources support unions:

- Tableau Desktop
- Excel
- Text file
- Google Sheets™
- JSON file (must have a .json, .txt, or .log extension)
- PDF file
- Amazon Redshift
- Aster® Database
- Cloudera™ Hadoop
- Google BigQuery™
- Hortonworks® Hadoop
- Vertica
- IBM® PureData™ (powered by Netezza®)
- Microsoft® SQL
- MySQL
- Oracle®
- PostgreSQL®
- Pivotal Greenplum®
- SAP® ASE™
- SAP® IQ™
- SAP® Teradata™

### Union and Databases

When working with database data sources, you should be aware of the following:

- Unions cannot be applied to stored procedures.
- Unions can be converted into custom SQL.

## Union Drop Area: Data Source Page Canvas

When creating unions from the **Data Source** page, you can select **New Union**, then select the tables or files to create a union. You can also select multiple files and drag them all to the **Union Drop** area in the canvas. If the canvas is empty, dragging multiple selected files or tables to the canvas will automatically create a union. If a table is already in the canvas, dragging additional tables over the existing table and dropping them onto it will automatically create a union.

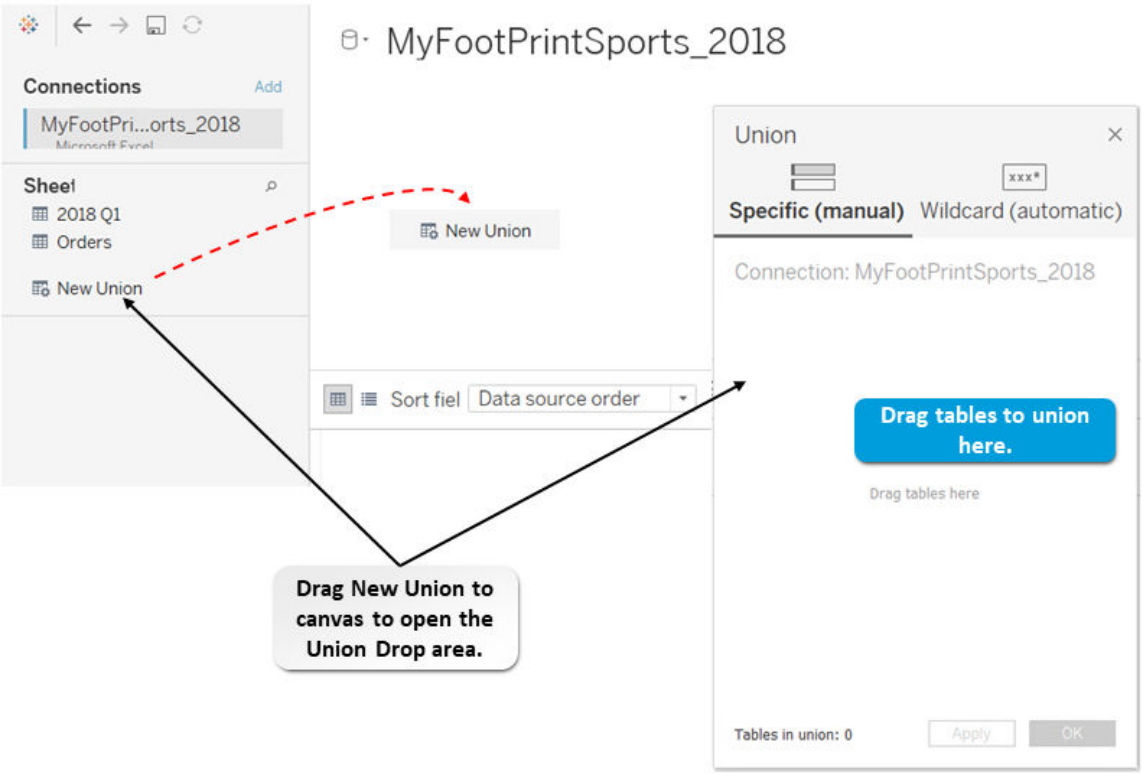



*Figure 2-11: Union Drop area: Data source page canvas.*

## Union with Wildcard Search

You can select the tables to include in a union, but if there are many potential tables or files you wish to include, you can use a wildcard search to automatically include tables. For example, if you wanted to find a character pattern in a series of worksheets in a spreadsheet, you could input the criteria to search tables based on the wildcard search.

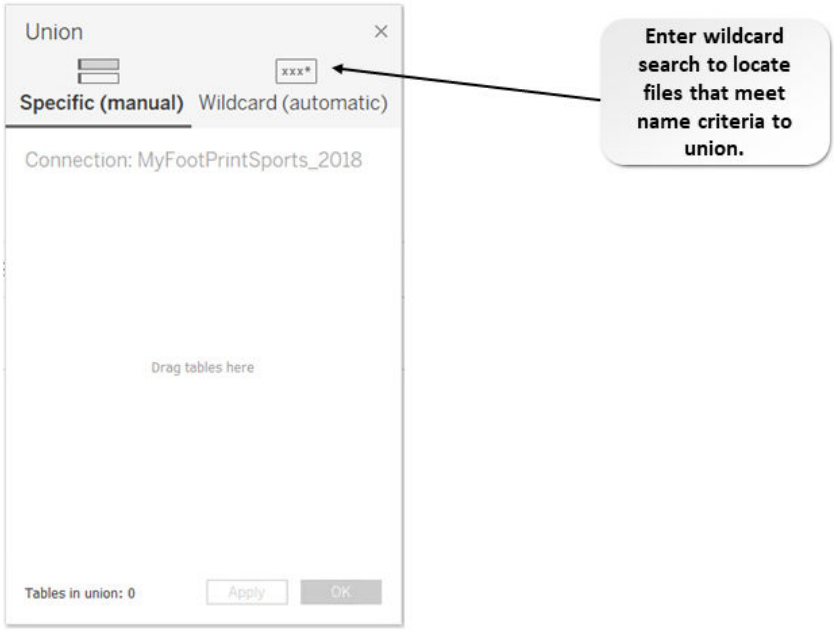



*Figure 2-12: A union with wildcard search.*

Wildcard searches allow you to automatically add the following:

- Tables in a database.
- Multiple sheets in a workbook.
- Multiple sheets in multiple workbooks in one or across multiple folders.
- Multiple text, JSON, and PDF files in one or across multiple folders.

All searches stay within the scope of the selected connection. For example, if you are searching for tables in a database, Tableau will search for tables in the database that is accessed via the selected connection. Results of the search are displayed in the left pane of the **Data Source** page. You can then modify, remove, and adjust which tables, sheets, and files to include in the union.

### Limitations

Wildcard search has the following limitations:

- When using a wildcard search to create unions of tables in PDF files, the search scope is limited to the pages initially scanned in the PDF.

## Mismatched Fields

Tableau's implementation of unions adds all rows to the updated table, including any duplicate values. Any fields named differently between the tables being joined in a union, or fields with the same names but different data types are considered mismatched. With mismatched fields, a union appends data in a similar way to an outer join. Mismatched rows are added, with null values added for columns that contain no data from other tables. The data grid can be used to merge fields and address some duplicate issues.

## Options for Addressing Mismatched Fields

In the data grid, you can merge fields to address duplicates and either remove null values or replace them with meaningful data. When you merge fields, a new field replaces the fields being merged. The first non-null value is used for each column where the row is mismatched.

For example, if sales data for multiple stores is in separate tables and you wish to get a view of sales company-wide, you can use unions to merge all store sales data into a single table. In this case, if databases were configured differently, any of the following issues might require you to perform extra steps to get clean, well-formed data.

| *Example* | *Resulting Issue* | *Potential Corrective Actions* |
|---|---|---|
| Northwest region used a field named Customer Address, and all other regions used the field named Address. | Null values in the Address field for customers in the Northwest region. An additional column named Customer Address that contains the address data for Northwest customers, and null values for all other regions' customers. | Merge the fields so that all data from the Customer Address field is placed into the Address field in the final combined table. |
| Northwest region used a numerical data type for the ZipCode field, while all other regions used a string type for ZipCode. | If fields have the same name but different data types, Tableau will change the name of one of the data fields to match others in a union. | If you don't like the resultant data type, right-click the Change Data Type icon and select Separate Inputs with Different Types to undo the merger. Then change the data types as required and perform the merger again. |

| *Example* | *Resulting Issue* | *Potential Corrective Actions* |
|---|---|---|
| Northwest region used a single field named Address to store customer's street address, city, state, and ZIP Code. All others used separate fields. | Tableau will combine address data from all sources into the Address field. The City, State, and ZIP Code fields for Northwest customers will contain null values. | Prior to performing a union, use the split function to break the Northwest Address field into new fields for City, State, and ZIP Code with names matching the tables from other regions. |

## Splitting a Field into Multiple Fields

For more information on splitting a field into multiple fields, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/split.html**.

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Merge and Append Data with Union.**

# ACTIVITY 2-4

## Merging and Appending Data with a Union

### Data File

C:\095205Data\Joining Data\MyFootPrintSports_2018.xlsx

### Before You Begin

Tableau Desktop is open.

### Scenario

You have been informed that data you receive in the future will be updates that only contain new data and none of the data you received previously. You have received a sample of what that data will look like so you can ensure it will work inside Tableau. You will create a union that appends the data update to an existing data source.

---

1. Review the data source.
   a) Open File Explorer and navigate to **C:\095205Data\Joining Data**.
   b) Open **MyFootPrintSports_2018.xlsx**.
   c) In Excel, on the **Orders** worksheet, review the data. This worksheet contains four years' worth of sales data from 2014 to 2017.

| ID | Order ID | Order Date | Ship Date | Shipping Method | Customer ID | Customer Name | City | State | Zip Code | Country | Region |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12037 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 12038 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 12039 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 12040 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 12041 | 2014-40922 | 1/14/2014 | 1/16/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 12045 | 2014-40923 | 1/15/2014 | 1/16/2014 | First Class | S-115901 | Albert Schmidt | Louisville | Kentucky | 40214 | United States | Southern U |
| 12051 | 2014-40927 | 1/19/2014 | 1/22/2014 | 2 Day | W-131951 | Denise Ward | Oakland | California | 94601 | United States | Western U |
| 12076 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 | Aaron Anderson | Lakewood | Ohio | 44107 | United States | Eastern US |
| 12077 | 2014-40935 | 1/27/2014 | 1/30/2014 | First Class | A-119651 | Aaron Anderson | Lakewood | Ohio | 44107 | United States | Eastern US |
| 12098 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 | Frances Burton | Austin | Texas | 78745 | United States | Central US |
| 12099 | 2014-40944 | 2/5/2014 | 2/9/2014 | Standard | B-132701 | Frances Burton | Austin | Texas | 78745 | United States | Central US |
| 12105 | 2014-40947 | 2/8/2014 | 2/11/2014 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 12108 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | Western U |
| 12109 | 2014-40951 | 2/12/2014 | 2/16/2014 | 2 Day | A-147251 | Judith Anderson | Salem | Oregon | 97301 | United States | Western U |

Orders | 2018 Q1

   d) In Excel, on the **2018 Q1** worksheet, review the data. This worksheet contains sales data for January through March of 2018.

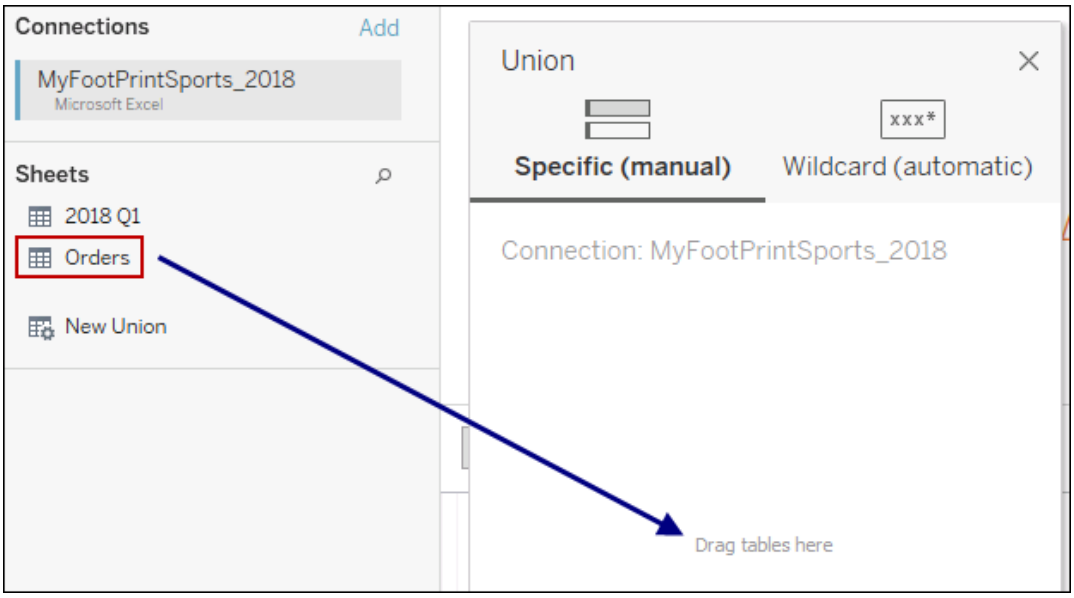
| ID | Order ID | Order Date | Ship Date | Ship Method | Customer ID | Customer Name | City | State | Zip Code | Country | Territory |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 22001 | 2018-42342 | 1/14/2018 | 1/16/2018 | 2 Day | Y-105551 | Julia Young | Los Angeles | California | 90049 | United States | Western US |
| 22002 | 2018-42342 | 1/14/2018 | 1/16/2018 | 2 Day | Y-105551 | Julia Young | Los Angeles | California | 90049 | United States | Western US |
| 22003 | 2018-42342 | 1/14/2018 | 1/16/2018 | 2 Day | Y-105551 | Julia Young | Los Angeles | California | 90049 | United States | Western US |
| 22004 | 2018-42342 | 1/14/2018 | 1/16/2018 | 2 Day | Y-105551 | Julia Young | Los Angeles | California | 90049 | United States | Western US |
| 22005 | 2018-42342 | 1/14/2018 | 1/16/2018 | 2 Day | Y-105551 | Julia Young | Los Angeles | California | 90049 | United States | Western US |
| 22006 | 2018-42343 | 1/15/2018 | 1/16/2018 | First Class | S-115901 | Albert Schmidt | Louisville | Kentucky | 40214 | United States | Southern U |
| 22007 | 2018-42347 | 1/19/2018 | 1/22/2018 | 2 Day | W-131951 | Denise Ward | Oakland | California | 94601 | United States | Western US |
| 22008 | 2018-42355 | 1/27/2018 | 1/30/2018 | First Class | A-119651 | Aaron Anderson | Lakewood | Ohio | 44107 | United States | Eastern US |
| 22009 | 2018-42355 | 1/27/2018 | 1/30/2018 | First Class | A-119651 | Aaron Anderson | Lakewood | Ohio | 44107 | United States | Eastern US |
| 22010 | 2018-42364 | 2/5/2018 | 2/9/2018 | Standard | B-132701 | Frances Burton | Austin | Texas | 78745 | United States | Central US |
| 22011 | 2018-42364 | 2/5/2018 | 2/9/2018 | Standard | B-132701 | Frances Burton | Austin | Texas | 78745 | United States | Central US |
| 22012 | 2018-42367 | 2/8/2018 | 2/11/2018 | 2 Day | S-122501 | Bruce Spencer | Green Bay | Wisconsin | 54302 | United States | Central US |
| 22013 | 2018-42371 | 2/12/2018 | 2/16/2018 | 2 Day | W-196301 | Alan Wilson | Dallas | Texas | 75220 | United States | Central US |
| 22014 | 2018-42371 | 2/12/2018 | 2/16/2018 | 2 Day | W-196301 | Alan Wilson | Dallas | Texas | 75220 | United States | Central US |

Orders | 2018 Q1

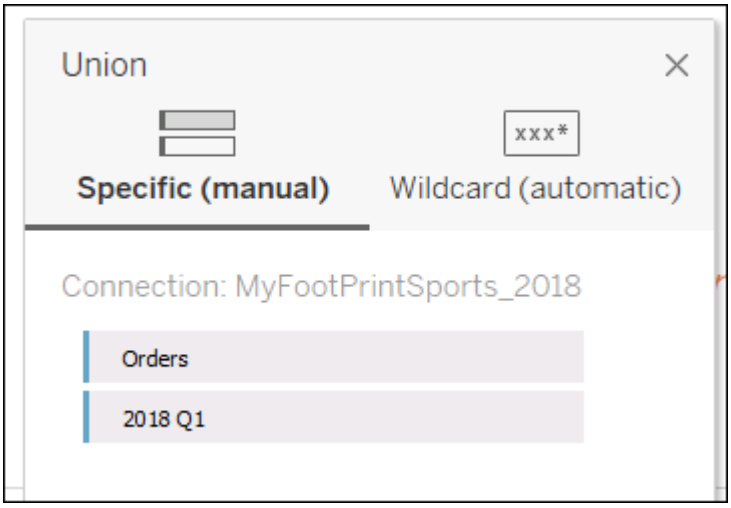
   e) Close Excel.
   f) If prompted to save changes, select **Don't Save**.

2. Connect to a new data source.
   a) In Tableau, at the bottom of the screen, select the **Data Source** tab.
   b) From the menu, select **Data→New Data Source**.
   c) Under **To a File**, select **Microsoft Excel**.
   d) In the **Open** dialog box, select **MyFootPrintSports_2018.xlsx** and then select **Open**.

3. Create a union.
   a) In the **Connections** pane, under **Sheets**, select and drag **New Union** to the canvas.
   b) In the **Connections** pane, under **Sheets**, select and drag **Orders** to the **Union** dialog box where it says **Drag tables here**.




   c) In the **Connections** pane, under **Sheets**, select and drag **2018 Q1** to the **Union** dialog box.




   d) Select **OK**.

4. Review the data union.

a) In the data grid, scroll down until you see **Order Date** entries for both **2018** and **2014**.

| # Union ID | Abc Union Order ID | Union Order Date | Union Ship Date |
|---|---|---|---|
| 22,050 | 2018-42417 | 3/29/2018 | 4/3/2018 |
| 22,051 | 2018-42418 | 3/30/2018 | 4/4/2018 |
| 22,052 | 2018-42418 | 3/30/2018 | 4/4/2018 |
| 12,037 | 2014-40922 | 1/14/2014 | 1/16/2014 |
| 12,038 | 2014-40922 | 1/14/2014 | 1/16/2014 |
| 12,039 | 2014-40922 | 1/14/2014 | 1/16/2014 |

b) Observe the **Shipping Method** column and how the **2018** dates have **null** as a value and the **2014** dates have data.

| Union Ship Date | Abc Union Shipping Method | Abc Union Customer ID |
|---|---|---|
| 4/3/2018 | *null* | A-147251 |
| 4/4/2018 | *null* | A-103151 |
| 4/4/2018 | *null* | A-103151 |
| 1/16/2014 | 2 Day | S-122501 |
| 1/16/2014 | 2 Day | S-122501 |
| 1/16/2014 | 2 Day | S-122501 |

c) Scroll to the right and observe that the **Region**, **Product Name**, and **Unit Price** columns also have null values for the 2018 and 2014 dates.

d) Continue to scroll to the right and observe there are columns for **Ship Method**, **Territory**, **Product**, and **Unit Cost** that have data for the 2018 dates and null values for the 2014 dates. These fields have different names in the 2018 data than the same fields in the 2014 to 2017 data.

| Abc Union Ship Method | Abc Union Territory | Abc Union Product | # Union Unit Cost |
|---|---|---|---|
| Standard | Western US | Women's Hiking Boot ... | 42.9900 |
| Standard | Central US | Basketball Shorts Wh... | 24.9900 |
| Standard | Central US | Baseball Glove Blue | 20.9900 |
| *null* | *null* | *null* | *null* |
| *null* | *null* | *null* | *null* |
| *null* | *null* | *null* | *null* |

## 5. Merge the shipping method fields.

a) Scroll back to the left and select the column header for the **Shipping Method** column to highlight it.

b) Scroll back to the right, hold down the **CTRL** key, and select the column header for the **Ship Method** column to highlight the column.

c) In the **Ship Method** column, select the **Down Arrow** and then select **Merge Mismatched Fields**.

d) Scroll back to the **Shipping Method & Ship Method** column and scroll down until you see **Order Date** entries for both **2018** and **2014**.

e) Observe that the null values are gone.

| Union Order Date | Union Ship Date | Abc Project Shipping Method ... |
|---|---|---|
| 3/29/2018 | 4/3/2018 | Standard |
| 3/30/2018 | 4/4/2018 | Standard |
| 3/30/2018 | 4/4/2018 | Standard |
| 1/14/2014 | 1/16/2014 | 2 Day |
| 1/14/2014 | 1/16/2014 | 2 Day |
| 1/14/2014 | 1/16/2014 | 2 Day |

f) In the **Shipping Method & Ship Method** column, select the **Down Arrow** and then select **Rename**.

g) Rename the column to ***Shipping Method***

## 6. Merge the remaining mismatched fields.

a) Select the column header for the **Region** column to highlight the column.

b) Locate the **Territory** column.

c) Hold down the **CTRL** key and select the column header for the **Territory** column to highlight the column.

d) In the **Territory** column, select the **Down Arrow** and then select **Merge Mismatched Fields**.
e) Locate the **Region & Territory** column and rename it to ***Region***
f) Repeat steps **a** through **e** to merge the **Product Name** and **Product** columns.
g) Rename the **Product Name & Product** column to ***Product Name***
h) Repeat steps **a** through **e** to merge the **Unit Price** and **Unit Cost** columns.
i) Rename the **Unit Price & Unit Cost** column to ***Unit Price***

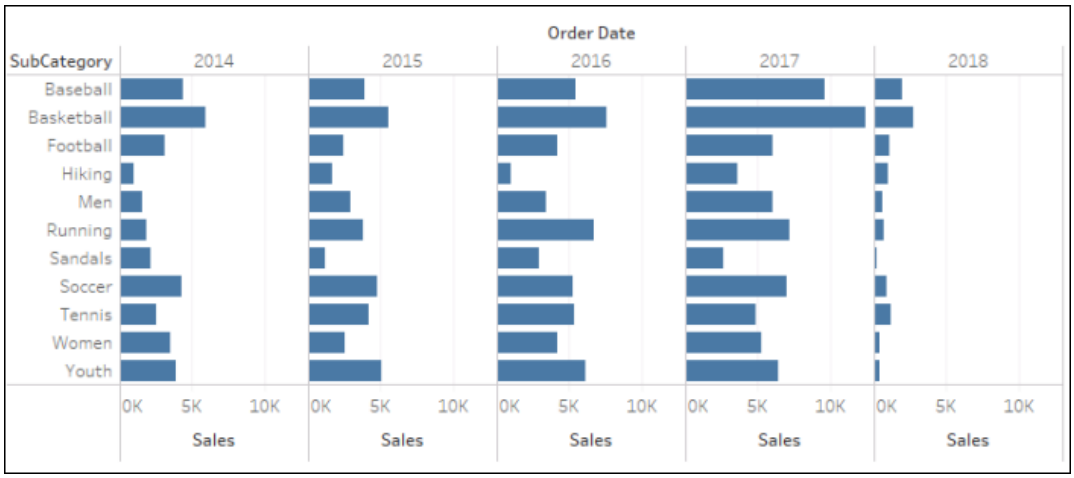
## 7. Clean up the fields.

a) At the bottom of the screen, select the **New Worksheet** button.

b) In the **Data** pane, verify the **Union** data source is selected.
c) In the **Data** pane, under **Dimensions**, observe the **Sheet** and **Table Name** fields. These fields contain which table and sheet a row of data came from in the data source. In this case, you don't want to know that, so you can hide the fields to remove clutter from the **Data** pane.
d) Right-click **Sheet** and then select **Hide**.
e) Right-click **Table Name** and then select **Hide**.

## 8. Create a view using the union data.

a) In the **Data** pane, under **Dimensions**, select and drag **Sub Category** to the **Rows** shelf.
b) In the **Data** pane, under **Dimensions**, select and drag **Order Date** to the **Columns** shelf.
c) In the **Data** pane, under **Measures**, select and drag **Sales** to the **Columns** shelf.
d) In the canvas, observe the view that shows sales data by subcategory for the years 2014 through 2017, plus the partial data for 2018.




## 9. Rename the **Sheet 6** worksheet.

a) Right-click the **Sheet 6** tab and select **Rename**.
b) Type ***Sales by Subcategory*** and press **Enter**.

## 10. Save the workbook.

a) From the menu, select **File→Save**.
b) From the menu, select **File→Close**.

# Summary

In this lesson, you joined data by configuring join types and join conditions. You did troubleshooting of issues related to joining data and fixed broken joins. Finally, you created unions with data to combine rows from multiple tables into a unified view.

### What types of joins will you create to combine data in your environment and why?

### In what scenarios might your organization use blends over joins?

| | |
|---|---|
| | **Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience. |

# 3 Accessing Data in PDFs

**Lesson Time: 1 hour, 30 minutes**

## Lesson Introduction

One of the most ubiquitous data sources in use are PDF files. Numerous reports are published in PDF format. Those reports often contain data that would be useful to data analysts. In this lesson, you will connect to PDFs to extract data from tables, and clean and organize that data for use in creating visualizations in Tableau®.

## Lesson Objectives

In this lesson, you will:

- Connect to PDFs.
- Clean and organize PDF data.

# TOPIC A

## Connect to PDFs

Since so many reports are distributed using the PDF file format, PDFs are often a great source of information for creating visualizations. In this topic, you will connect to PDFs to extract data.

### PDF Data Source Connections

Many documents and reports are published and distributed in PDF format. These documents may contain data that is valuable or necessary for data analysis or reporting. For example, you may wish to compare information published by leading research centers for industries you serve to your own data to find trends and gaps. In another case, a stock brokerage may wish to take statistical data about stock prices from a published PDF. In these cases it's helpful to be able to pull data from a PDF file into Tableau.

Tableau can connect to PDF using the PDF connector. Specifically, the PDF connector allows Tableau to bring crosstab data tables in PDF files that have defined rows and columns and a single data row for each entry in the table into Tableau. Since PDF tables don't have any metadata with them, there may be additional post-connection cleanup required.

### Connections to PDF Tables

You can connect to a PDF file by selecting the PDF file type from the **Connect** pane when adding a data source. You can then enter or browse to the location of the PDF file you're looking for and select it to begin the connection process. When you select a file, a dialog box appears and gives you the following options for locating tables in the file:

- **Scan**: To scan the PDF file and enumerate all tables on all pages.
- **Single page**: To scan a single page and enumerate the tables on that page.
- **Range of pages**: To scan a range of pages, such as 10 to 20, and enumerate the tables on those pages.

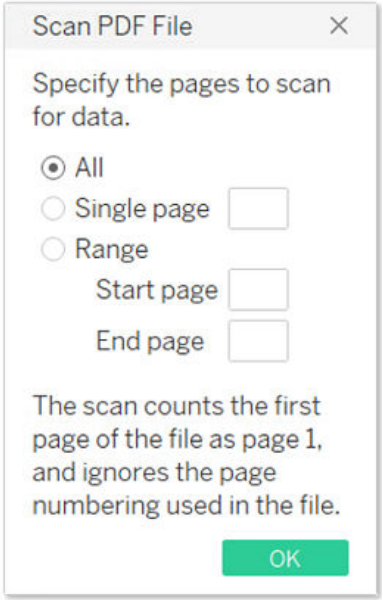



*Figure 3-1: The Scan PDF file dialog box.*

Each page that has a viable table that Tableau can connect to will be listed on the left pane of the **Data Source** page. You can drag tables you're interested in onto the canvas to see the data in the data grid. If headers or content is not formatted as expected, you may need to adjust the data interpreted to format it in the way you intended.

If you wish to change the pages that were scanned, you can select **Rescan PDF file** under the data connection drop-down. You can then change the type of scan, or select a new page or pages to scan.

> **Note:** Tableau looks at absolute page numbers in the document, starting by counting the first page as page 1, and counting from there. Those page numbers may not line up with the page numbers printed in the PDF, as some PDFs start with numbering the front matter in Roman numerals. In a case where a PDF has 10 pages of front matter and a table on a page with page number 10 in the PDF, that same table will be on page 20 using absolute numbering.

## Merge Tables and Span Multiple Pages with Unions

In some cases, tables in PDFs may span multiple pages. Since tables are technically part of separate pages in the PDF, the scan will identify and list them as separate tables on separate pages. You can combine these tables using unions. You can do this by selecting to create a **New Union**, or by selecting all the tables to union, and dragging them to the **Union Drop** area in the canvas. You can use the Table Name column in the data grid to verify the union has correctly brought all the tables into the union.

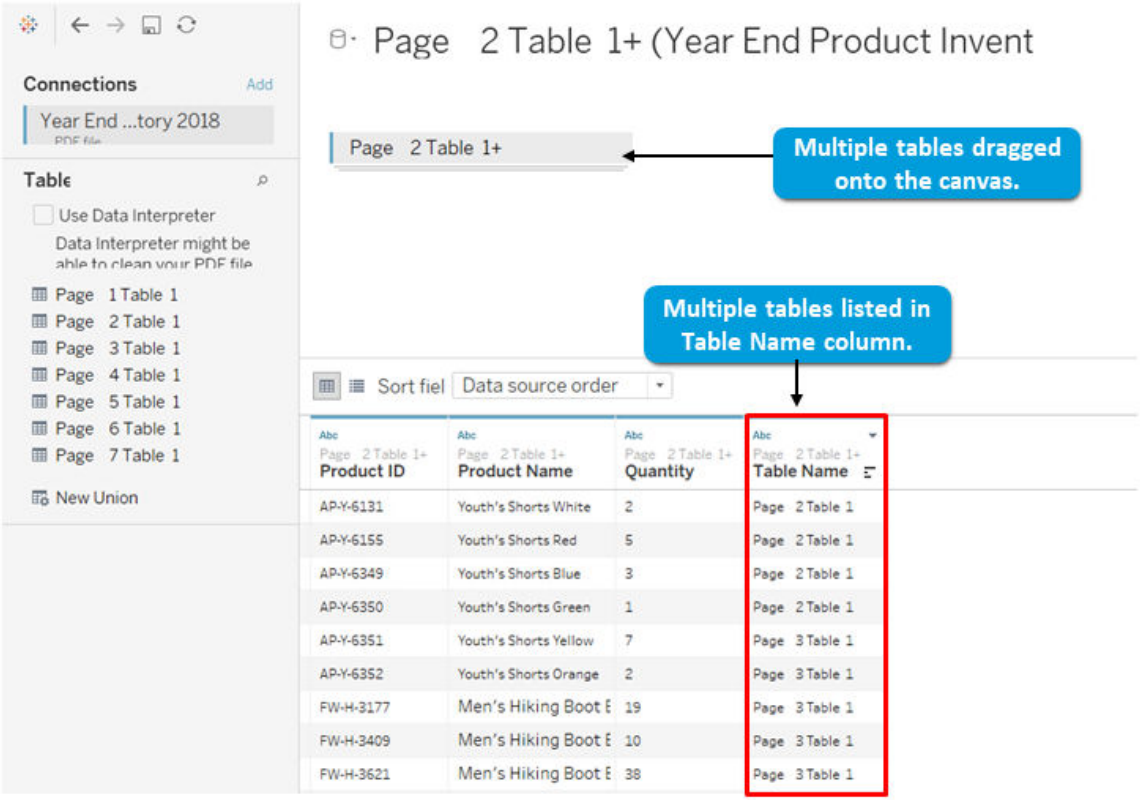



*Figure 3-2: Union connections in a PDF.*

## Tables with Complex Formatting

Tables that are cross tabs with simple, well-defined columns and rows work very well with the PDF connector. However, in many cases, tables may have nested headers, established hierarchies, multiple rows that need to be interpreted as a single row, and even nested sub-tables. Even colors and shading used in the PDF can change how data is interpreted by the connector. The PDF

connector can still connect and bring data from these tables into Tableau, yet it will likely take more effort to clean up the data and get it the way you want for analysis and reporting in Tableau. For this reason, it's important to have a strong understanding of the data you're trying to import and what you expect to get.

The types of issues you may encounter include:

- **The same table being detected by the PDF connector as multiple tables**. In this case, multiple tables from the same page will be listed in the left pane.
- **Multiple columns combined into a single column.** In this case, you could perform a ***custom split*** to ***split*** the column on a common separator and then rename the split columns.
- **Null values**: If a table contains sub-headers and Tableau interprets them as rows, there will be "no" in the columns for that row, leaving a row of null values. It's also possible that a single row will be read as two different rows by Tableau. In this case, both rows will have an incorrect first field and one row will have all null values.
- **Totals included as data.** In some cases, fields that are totals may be included as data. These should be filtered out.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Connect to PDF Data Sources.** |

# ACTIVITY 3-1
## Connecting to PDF Data Sources

### Data Files

C:\095205Data\Accessing Data in PDFs\Year End Product Inventory 2018.pdf

C:\095205Data\Accessing Data in PDFs\MyFootPrintSports_2018_3.twb

### Before You Begin

Tableau Desktop is open.

### Scenario

You have received a PDF from a year-end inventory that includes the in-stock quantities for all products. You want to be able to access the inventory data from Tableau so you can create a view to better visualize your in-stock quantities.

---

1. Open the worksheet for the lesson.
   a) From the menu, select **File→Open**.
   b) In the **Open** dialog box, browse to the **C:\095205Data\Accessing Data in PDFs** folder.
   c) Select **MyFootPrintSports_2018_3** and select **Open**.

2. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Accessing Data in PDFs** folder.
   c) In the **File name** box, type ***MyCopy_MyFootPrintSports_2018_3*** and select **Save**.

3. Review the PDF source file.
   a) Open File Explorer and navigate to **C:\095205Data\Accessing Data in PDFs**.
   b) Open **Year End Product Inventory 2018.pdf**.

c) In your PDF reader, review the data. This PDF contains a simple table that lists in-stock product quantities from a year-end inventory. It is a simple table that should be easy to bring into Tableau.

**Product Inventory 2018**

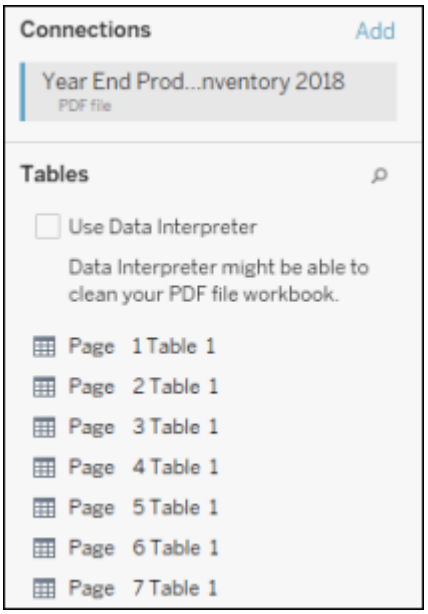
| Product ID | Product Name | Quantity |
|---|---|---|
| AP-M-3124 | Men's Tank Top Black | 32 |
| AP-M-3125 | Men's Tank Top White | 17 |
| AP-M-3373 | Men's Tank Top Red | 40 |
| AP-M-3407 | Men's Tank Top Blue | 16 |
| AP-M-3748 | Men's Tank Top Green | 48 |
| AP-M-4140 | Men's Tank Top Yellow | 13 |
| AP-M-4142 | Men's Tank Top Orange | 27 |
| AP-M-4300 | Men's T-Shirt Black | 23 |
| AP-M-4788 | Men's T-Shirt White | 9 |
| AP-M-4799 | Men's T-Shirt Red | 14 |
| AP-M-4870 | Men's T-Shirt Blue | 37 |
| AP-M-4871 | Men's T-Shirt Green | 27 |
| AP-M-4872 | Men's T-Shirt Yellow | 16 |
| AP-M-4873 | Men's T-Shirt Orange | 22 |
| AP-M-4875 | Men's Pants Black | 11 |
| AP-M-4876 | Men's Pants White | 59 |
| AP-M-4877 | Men's Pants Red | 14 |

d) Close your PDF reader.

## 4. Connect to a simple PDF.

a) In Tableau, at the bottom of the screen, select the **Data Source** tab.
b) From the menu, select **Data→New Data Source**.
c) Under **To a File**, select **PDF file**.
d) In the **Open** dialog box, navigate to **C:\095205Data\Accessing Data in PDFs**.
e) Select **Year End Product Inventory 2018.pdf** and then select **Open**.
f) In the **Scan PDF File** dialog box, review the options for which pages you can select.
g) Select **OK** with **All** selected.

h) In the left pane, under **Tables**, observe the list of seven pages of tables. Based on the names of the tables, Tableau recognized that it might be a single table that spans multiple pages.




i) In the left pane, under **Tables**, select and drag **Page 1 Table 1** to the canvas.

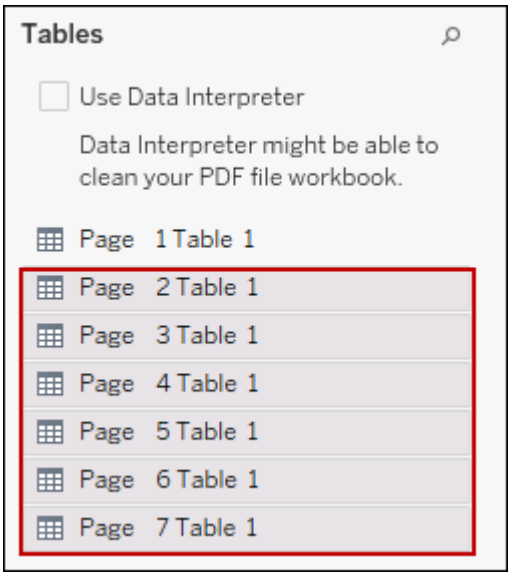
j) In the grid, review the data rows and columns. Observe that there are no null values and the values are consistent per column.

| Abc<br>Page 1 Table 1<br>**Product ID** | Abc<br>Page 1 Table 1<br>**Product Name** | Abc<br>Page 1 Table 1<br>**Quantity** |
|---|---|---|
| AP-M-3124 | Men's Tank Top Black | 32 |
| AP-M-3125 | Men's Tank Top White | 17 |
| AP-M-3373 | Men's Tank Top Red | 40 |
| AP-M-3407 | Men's Tank Top Blue | 16 |
| AP-M-3748 | Men's Tank Top Green | 48 |

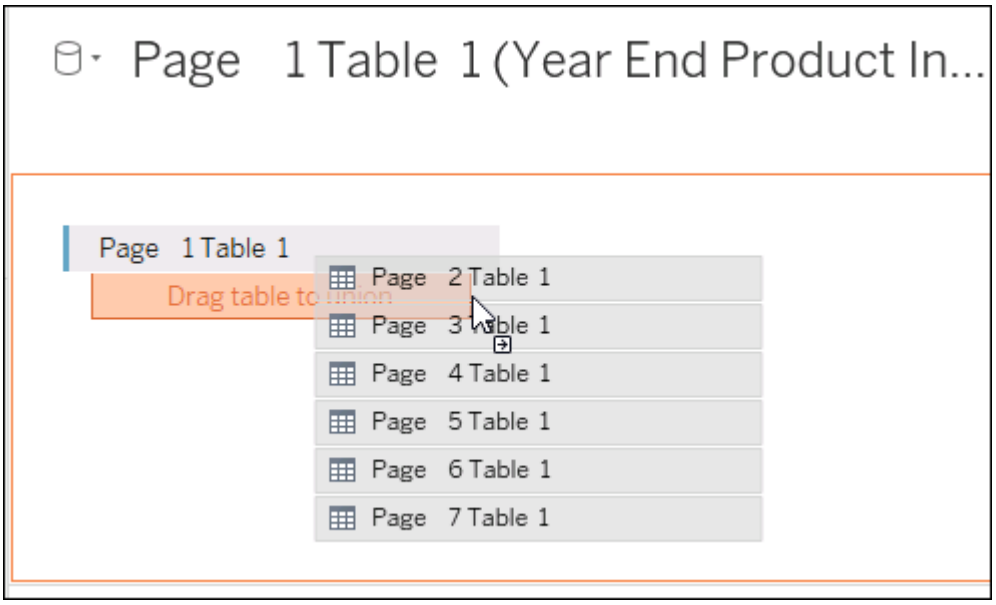
## 5. Add additional PDF pages.

a) In the left pane, under **Tables**, select **Page 2 Table 1**.

b) Hold down the **Shift** key and select **Page 7 Table 1** to highlight those tables.




c) In the left pane, under **Tables**, select and drag the selected tables to the canvas, hover over **Page 1 Table 1** until the **Drag table to union** box appears and is filled with an orange color, and then release them.

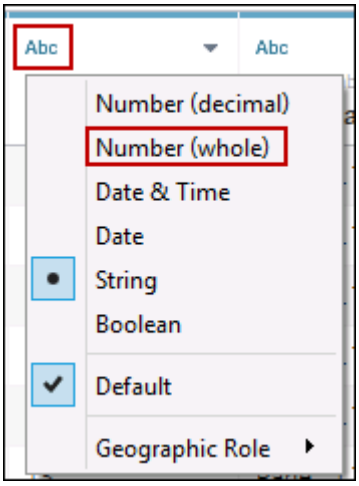


d) In the grid, scroll down and review the data rows and columns. Observe the **Table Name** column and verify the data from the different tables is present.

| Abc<br>Page 1 Table 1+<br>**Product ID** | Abc<br>Page 1 Table 1+<br>**Product Name** | Abc<br>Page 1 Table 1+<br>**Quantity** | Abc<br>Page 1 Table 1+<br>**Table Name** |
|---|---|---|---|
| AP-W-3161 | Women's T-Shirt Red | 32 | Page 1 Table 1 |
| AP-W-3162 | Women's T-Shirt Blue | 9 | Page 1 Table 1 |
| AP-W-3163 | Women's T-Shirt Green | 49 | Page 1 Table 1 |
| AP-W-3164 | Women's T-Shirt Yellow | 13 | Page 2 Table 1 |
| AP-W-3165 | Women's T-Shirt Oran... | 5 | Page 2 Table 1 |
| AP-W-3169 | Women's Pants Black | 17 | Page 2 Table 1 |

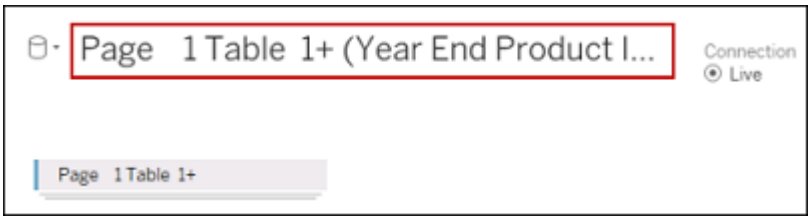
## 6. Clean up the data connection.

a) In the **Quantity** column header, select the **Data Type** icon Abc and then select **Number (whole)**.




b) In the **Table Name** column header, select the drop-down arrow, and then select **Hide**.

c) At the top of the canvas, select **Page 1 Table 1+ (Year End Product Inventory 2018** to highlight the text.



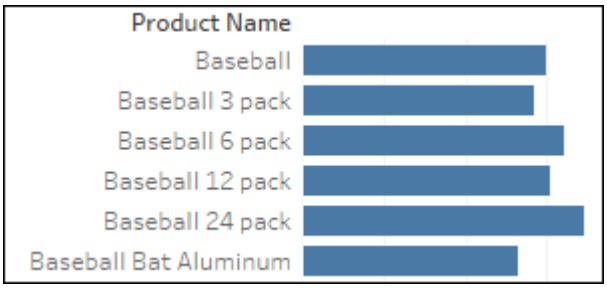

d) Type ***Product Inventory*** and press **Enter** to rename it.

e) At the bottom of the screen, select the **New Worksheet** button.

f) In the **Data** pane, verify the **Product Inventory** data source is selected.

g) In the **Data** pane, under **Dimensions**, select and drag **Quantity** to the **Measures** section.
**Quantity** is actually a measure, so you will move it to the **Measures** section to avoid confusion later.

7. Create a simple view for the **Product Inventory** data source.
   a) In the **Data** pane, under **Dimensions**, select and drag **Product Name** to the **Rows** shelf.
   b) In the **Data** pane, under **Measures**, select and drag **Quantity** to the **Columns** shelf.
   c) In the canvas, observe the product quantities from the last inventory.




   d) Right-click the **Sheet 7** tab and select **Rename**.
   e) Type ***Product Inventory*** and press **Enter**.

# TOPIC B

## Clean Up and Organize PDF Data

Much like other data sources, connecting to PDF data can cause errors and other issues in the data. In this topic, you will clean and organize PDF data for use in visualizations.

### Data Source Filters

If you don't need all the data being brought in from a data source, you can add a ***data source filter*** from the upper-right corner of the **Data Source** page. This restricts what data is available for use in creating views and analysis. There are many reasons for using data source filters, including:

- Restricting data users can see for security or privacy.
- Removing unnecessary data by showing only the subset of data necessary for exploration, analysis, or reporting.
- To improve performance with large data sets by reducing the amount of data processed by Tableau.

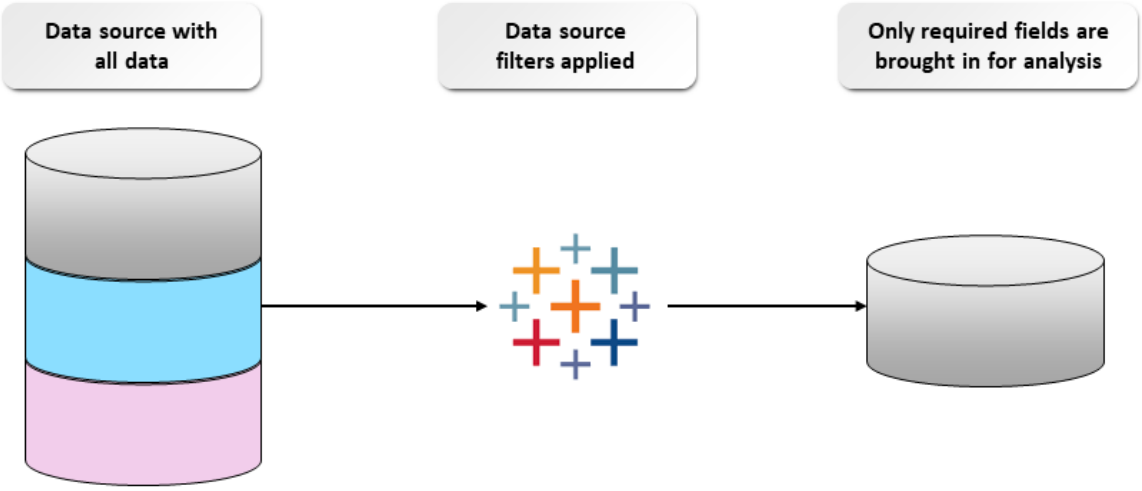



*Figure 3-3: Data source filters.*

#### Additional Information

For more information, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/filtering_datasource.html**.

### Data Cleanup with Data Source Filters

Data source filters can remove unwanted data from your data source. For example, if you have rows that contain null values, one way to get rid of them is to add a data source filter that excludes any row with a column that contains null values. You must be certain this very broad filter will not remove data you wish to keep. You can do the same to move rows that contain totaled data. In that case, you might have to manually select those rows.

### Finalizing Data Organization

To finalize the data and make it ready to work with and build visualizations from it, you may need to do the following.

| Task | Description |
|---|---|
| ***Pivot*** data | Create pivots of data to organize columns based on the data you wish to explore and analyze. |
| Adjust data types | Change the data types for any fields that require it, such as changing fields assigned string values to dates of date fields, or changing to a more relevant numeric data type. |
| Re-alias field members | ***Aliases*** in Tableau allow you to assign more meaningful names to members of a field (the data in the field) to be displayed in the Viz. Using aliases preserves the data from the source as-is, but shows the more meaningful name in the Viz. You may need to re-alias if Tableau had difficulty parsing a complex table that generated split rows or other anomalies with member data. Re-aliasing tends to be a manual process that requires referencing the original data source. |
| Perform additional filtering | You may wish to apply additional data source filters as you get the data more organized to remove specific members, rows, columns, or nulls. |
| Adjust dimensions and measures | Once you open a worksheet based on the new PDF data source, you may find some measures or dimensions incorrectly identified. Review the **Data** pane and drag any dimensions that are really measures to the measures area, and vice versa. |

## Group and Hierarchy Recreation

Complex, nested groupings in tables in a PDF is one of the most common reasons the PDF connector has issues that must be cleaned up. Once data is cleaned up, it may be desirable to re-create the ***groups*** in the visualization for clarity or to aid in analysis. This can often be done in the **Data** pane or in the Viz the way you'd create any group, which is by selecting members that should be part of the group, and right-clicking to select the **Group** icon to group them. You can also rename groups and create groups that have only one member. From there, you can create ***hierarchies*** from the groups to allow users to drill into items in the Viz to get more detail.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Clean Up and Organize PDF Data.** |

# ACTIVITY 3-2

## Cleaning Up and Organizing PDF Data

### Data File

C:\095205Data\Accessing Data in PDFs\Profit by Category 2017.pdf

### Before You Begin

The **MyCopy_MyFootPrintSports_2018_3** workbook is open in Tableau Desktop.

### Scenario

You have received a PDF containing profit percentage data for different categories and subcategories that you want to be able to access from Tableau. The table in this PDF is not as simple as the previous one, so you will have to clean up the data before it can be used.

---

1. Review the PDF source file.
   a) In File Explorer, open **Profit by Category 2017.pdf**.
   b) In your PDF reader, review the data. This PDF contains profit percentages for different product categories by month. The more complicated layout of this table will require more effort to bring into Tableau.

**Profit for the year 2017 for all categories and subcategories.**

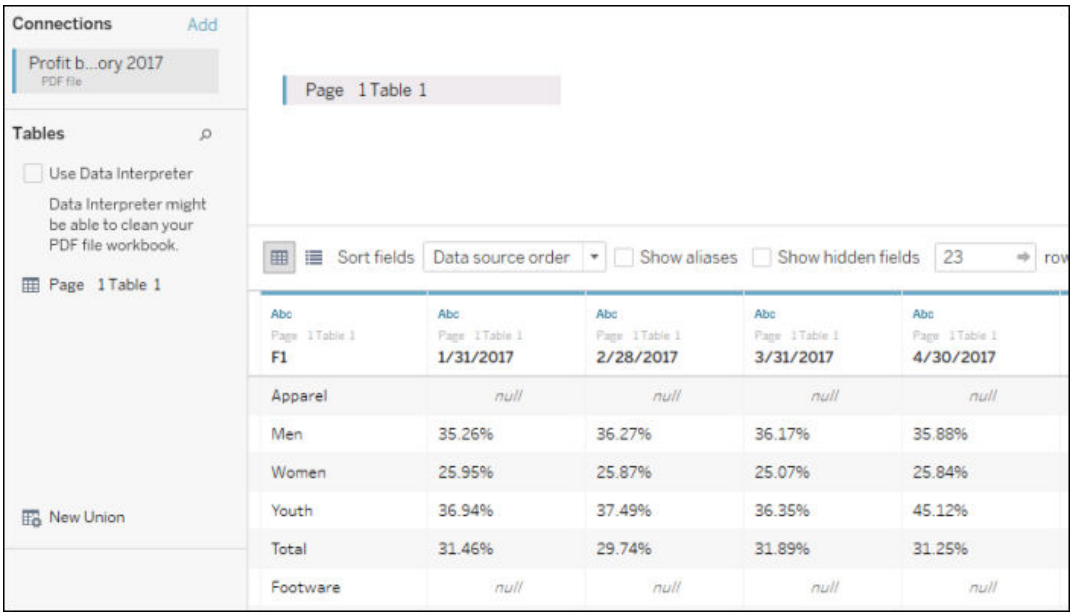
| | Year 2017 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1/31/2017 | 2/28/2017 | 3/31/2017 | 4/30/2017 | 5/31/2017 | 6/30/2017 | 7/31/2017 | 8/31/2017 | 9/30/2017 | 10/31/2017 | 11/30/2017 | 12/31/2017 |
| Apparel | | | | | | | | | | | | |
| Men | 35.26% | 36.27% | 36.17% | 35.88% | 36.93% | 36.29% | 36.04% | 37.04% | 36.60% | 36.27% | 35.43% | 36.17% |
| Women | 25.95% | 25.87% | 25.07% | 25.84% | 26.94% | 25.38% | 26.11% | 27.10% | 26.87% | 25.63% | 26.34% | 26.63% |
| Youth | 36.94% | 37.49% | 36.35% | 45.12% | 38.00% | 43.74% | 40.47% | 43.68% | 44.44% | 44.00% | 44.04% | 45.29% |
| Total | 31.46% | 29.74% | 31.89% | 31.25% | 31.30% | 32.56% | 31.19% | 31.36% | 31.81% | 30.89% | 31.39% | 31.75% |
| Footware | | | | | | | | | | | | |
| Hiking shoes and boots | 27.15% | 27.80% | 27.38% | 28.48% | 28.55% | 28.57% | 27.80% | 28.29% | 28.09% | 26.83% | 27.34% | 28.27% |
| Running | 39.44% | 39.44% | 39.16% | 38.44% | 38.97% | 39.07% | 38.31% | 38.93% | 38.58% | 38.42% | 39.21% | 39.13% |
| Sandals | 43.26% | 40.10% | 46.76% | 44.13% | 46.54% | 45.66% | 45.87% | 44.84% | 45.57% | 46.37% | 44.50% | 44.61% |
| Youth shoes and boots | 58.11% | 57.10% | 57.55% | 57.00% | 57.70% | 56.65% | 56.43% | 56.85% | 57.11% | 56.73% | 57.26% | 57.44% |
| Total | 48.54% | 47.43% | 44.03% | 47.88% | 46.93% | 41.35% | 46.21% | 42.36% | 46.08% | 46.60% | 44.85% | 44.72% |
| Sports | | | | | | | | | | | | |
| Baseball gear | 21.28% | 23.08% | 21.16% | 20.94% | 21.46% | 22.02% | 22.38% | 21.40% | 21.10% | 21.42% | 21.84% | 21.85% |
| Basketball gear | 36.56% | 35.10% | 37.04% | 36.71% | 35.66% | 36.24% | 36.19% | 37.17% | 36.71% | 35.93% | 37.11% | 36.98% |
| Football gear | 25.86% | 27.68% | 26.88% | 26.62% | 26.05% | 26.54% | 26.05% | 26.60% | 26.95% | 26.59% | 26.29% | 26.80% |
| Soccer gear | 46.57% | 45.83% | 47.29% | 44.62% | 46.24% | 46.13% | 45.97% | 46.10% | 46.61% | 46.03% | 46.70% | 45.93% |
| Tennis gear | 17.51% | 15.76% | 16.65% | 17.66% | 17.33% | 17.13% | 17.33% | 17.78% | 16.24% | 16.79% | 17.25% | 17.38% |
| Total | 30.77% | 29.60% | 31.78% | 29.09% | 30.19% | 30.62% | 29.55% | 29.96% | 31.68% | 30.48% | 31.92% | 29.95% |
| Grand Total | 35.59% | 35.03% | 34.15% | 34.74% | 36.18% | 34.07% | 33.28% | 33.90% | 35.49% | 35.50% | 35.97% | 34.16% |

These are raw profit percentages based on sales and costs for the year 2017.

   c) Close your PDF reader.

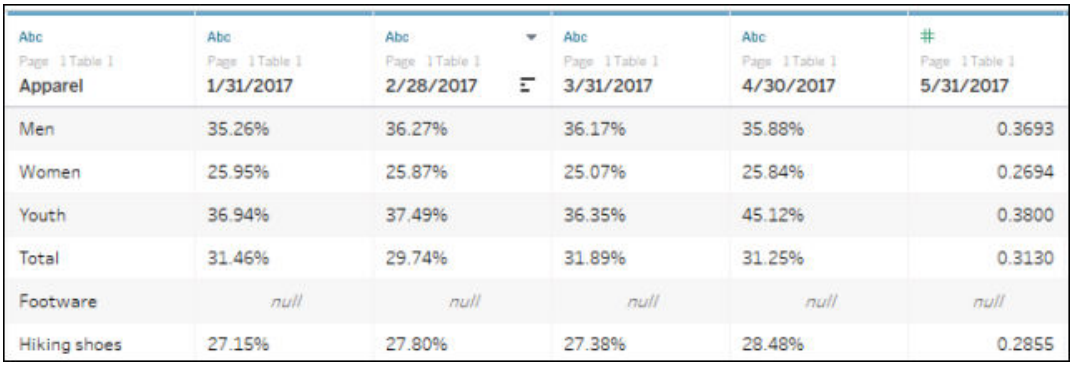
2. Connect to the complicated PDF.
   a) In Tableau, at the bottom of the screen, select the **Data Source** tab.
   b) From the menu, select **Data→New Data Source**.
   c) Under **To a File**, select **PDF file**.
   d) In the **Open** dialog box, select **Profit by Category 2017.pdf** and then select **Open**.
   e) In the **Scan PDF File** dialog box, select **OK** with **All** selected.

f) Observe that Tableau recognized one table, but that there are null values, a missing header, and other potential issues.




g) In the left pane, under **Tables**, check the **Use Data Interpreter** check box.

h) In the grid, observe that **Data Interpreter** removed the first row of null values, added the header (even though it is incorrect), and changed the data type for the last eight columns to a number.




3. Review the unwanted rows.

a) In the grid, scroll down through the rows and observe that five of the rows with null values are either category headers (**Footware** and **Sports**) or subcategory headers that had wrapped to the second line in the PDF (**and boots** and **gear**).

| Abc<br>Page 1 Table 1<br>Apparel | Abc<br>Page 1 Table 1<br>1/31/2017 | Abc<br>Page 1 Table 1<br>2/28/2017 |
|---|---|---|
| Total | 31.46% | 29.74% |
| Footware | *null* | *null* |
| Hiking shoes | 27.15% | 27.80% |
| and boots | *null* | *null* |

b) In the grid, scroll down through the rows and observe that there are three **Total** rows and one **Grand Total** row.

| | | |
|---|---|---|
| Tennis gear | 17.51% | 15.76% |
| Total | 30.77% | 29.60% |
| Grand Total | 35.59% | 35.03% |

c) In the grid, scroll down to the last row and observe that it is a sentence that has been split up over the first six columns.

| | | | | | |
|---|---|---|---|---|---|
| Grand Total | 35.59% | 35.03% | 34.15% | 34.74% | 0.3618 |
| These are raw profit | percentages | based on sal | es and costs f | or the year 2 | 17.0000 |

## 4. Filter out the unwanted rows.

a) In the upper-right corner of the canvas, under **Filters**, select **Add**.
b) In the **Edit Data Source Filters** dialog box, select **Add**.
c) In the **Add Filter** dialog box, select **Apparel** and select **OK**.
The rows you will filter first have values in the first column that you will filter on.
d) In the **Filter [Apparel]** dialog box, check the check boxes for **and boots**, **Footware**, **gear**, **Grand Total**, **Sports**, **These are raw profit**, and **Total**.
e) Under the list, check the **Exclude** check box.
f) Select **OK** twice.
g) In the grid, observe that all of the unwanted rows have been removed.

## 5. Pivot the data.

a) Select the **1/31/2017** column header.
b) Hold down the **Shift** key and select the **12/31/2017** column header.
c) In the **12/31/2017** date column, select the drop-down arrow, and then select **Pivot**.

d) In the grid, review the results. There is now a column for the dates and a column for the profit percent.

| Abc<br>Page 1 Table 1<br>Apparel | Abc<br>Pivot<br>Pivot Field Names | Abc<br>Pivot<br>Pivot Field Values |
|---|---|---|
| Men | 1/31/2017 | 35.26% |
| Women | 1/31/2017 | 25.95% |
| Youth | 1/31/2017 | 36.94% |

6. Edit aliases for subcategories that had wrapped text in the PDF.
   a) In the **Apparel** column header, select the drop-down arrow and select **Aliases**.
   b) In the **Edit Aliases [Apparel]** dialog box, select the **Basketball** row.
   c) In the **Value (Alias)** column, type ***Basketball gear***
   d) For **Hiking shoes**, change the **Value (Alias)** to ***Hiking shoes and boots***
   e) For **Youth shoes**, change the **Value (Alias)** to ***Youth shoes and boots***
   f) Select **OK**.

7. Rename the fields.
   a) In the **Pivot Field Names** column header, select the drop-down arrow and select **Rename**.

   > **Note:** Note there are two separate fields created by this pivot. The first is named **Pivot Field Names** and the second is **Pivot Field Values**. Make sure you are renaming the correct field.

   b) For the column name, type ***Date*** and press **Enter**.
   c) Rename the **Pivot Field Values** column header to ***Profit Percent***
   d) Rename the **Apparel** column header to **Subcategory**.
   e) At the top of the canvas, select **Page 1 Table 1 (Profit by Category 2017)** to highlight the text.
   f) Type ***Profit by Category*** and press **Enter** to rename it.

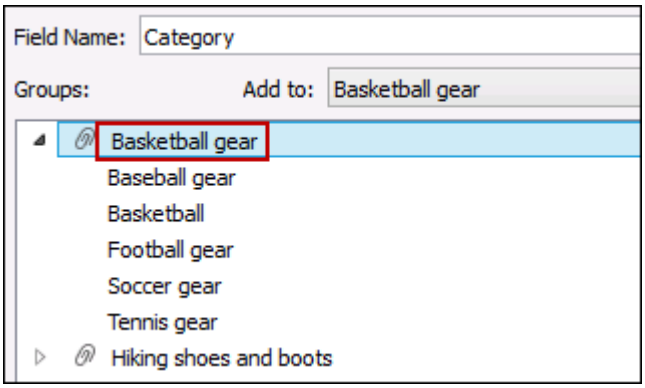
8. Change the data types.
   a) For **Date**, select the **Data Type** icon Abc and then select **Date**.
   b) For **Profit Percent**, select the **Data Type** icon Abc and then select **Number (decimal)**.

9. Configure the settings for the **Profit Percent** field.
   a) At the bottom of the screen, select the **New Worksheet** button.
   b) In the **Data** pane, verify the **Profit by Category** data source is selected.
   c) In the **Data** pane, under **Dimensions**, select and drag **Profit Percent** to the **Measures** section.

      **Profit Percent** is actually a measure, so you will move it to the **Measures** section to avoid confusion later.
   d) In the **Data** pane, under **Measures**, right-click **Profit Percent** and select **Default Properties→Number Format**.
   e) In the **Default Number Format [Profit Percent]** dialog box, select **Percentage**.
   f) Select **OK**.
   g) In the **Data** pane, under **Measures**, right-click **Profit Percent** and select **Default Properties→Aggregation→Average**.

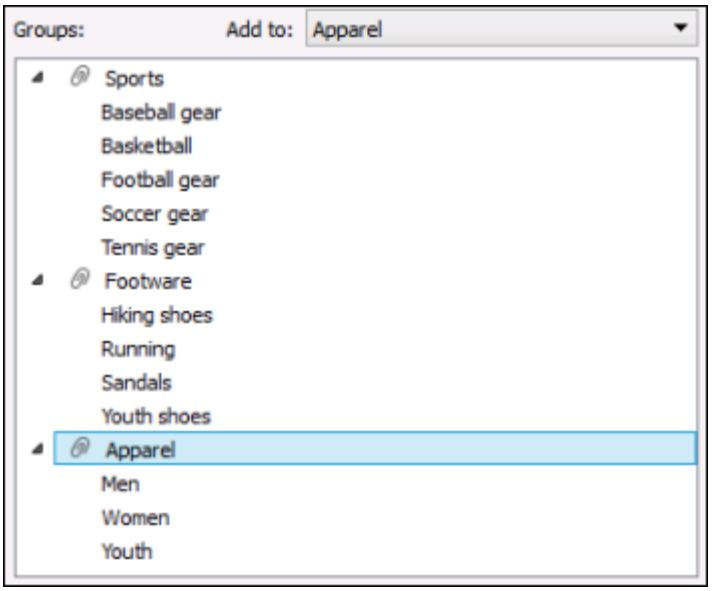
## 10. Replicate the main category groups.

a) In the **Data** pane, under **Dimensions**, right-click **Subcategory** and select **Create→Group**.
b) In the **Create Group [Subcategory]** dialog box, in the **Field Name** box, type ***Category***
c) In the **Groups** list, select **Baseball gear**.
d) Hold down the **CTRL** key and select **Basketball gear**, **Football gear**, **Soccer gear**, and **Tennis gear**.
e) Select **Group**.
f) In the **Groups** list, right-click the **Basketball gear** group and select **Rename**.




**Note:** The aliases you created earlier in this activity appear in the view, but when creating groups, you see the actual field name.

g) Type ***Sports*** and press **Enter**.
h) Repeat steps **c** through **g** to group **Hiking shoes and boots**, **Running**, **Sandals**, and **Youth shoes and boots** into a group named **Footware**.
i) Repeat steps **c** through **g** to group **Men**, **Women**, and **Youth** into a group named **Apparel**.



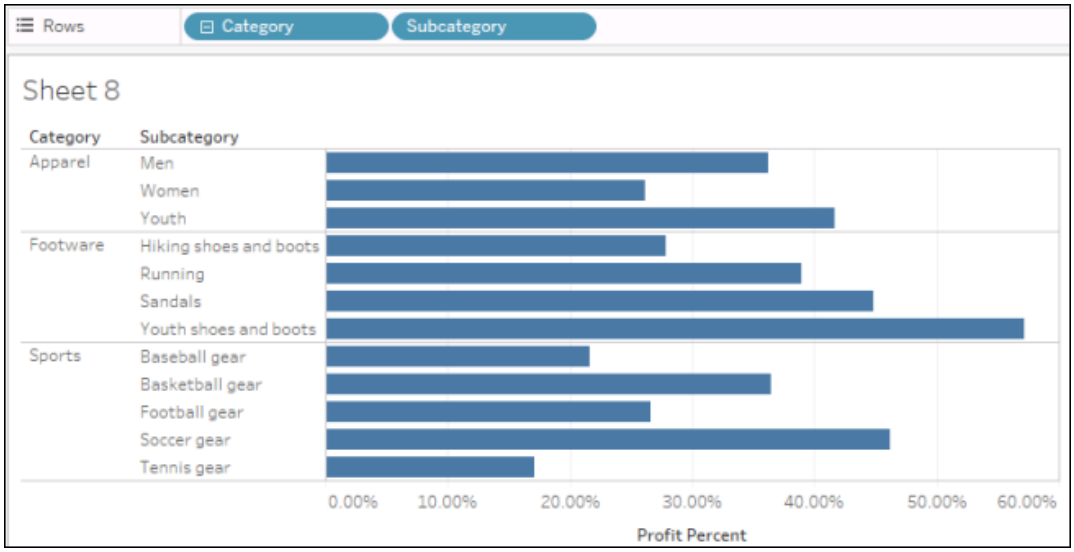

j) Select **OK**.

## 11. Create a hierarchy for **Category** and **Subcategory**.

a) In the **Data** pane, select and drag **Subcategory** onto **Category**.
b) In the **Create Hierarchy** dialog box, in the **Name** box, type ***Products***
c) Select **OK**.

12. Create a simple view for the **Profit by Category** data source.

a) In the **Data** pane, under **Dimensions**, select and drag **Products** to the **Rows** shelf.
b) In the **Data** pane, under **Measures**, select and drag **Profit Percent** to the **Columns** shelf.
c) In the **Rows** shelf, in the **Category** pill, select the plus sign to expand the field.
d) In the canvas, observe the profit percentages by category and subcategory.




e) Right-click the **Sheet 8** tab and select **Rename**.
f) Type ***Profit by Category*** and press **Enter**.

13. Save the workbook.

a) From the menu, select **File→Save**.
b) From the menu, select **File→Close**.

# Summary

In this lesson, you connected to PDFs to extract table data from them to visualize in Tableau. You unioned tables, spanning multiple pages to create a combined data set, and addressed complex formatting issues. You applied data source filters and grouped data to clean up, organize, and finalize PDF data brought into Tableau.

### How much data do you have in PDF format that you might want to visualize in Tableau?

### How complex is the table data in those PDFs?

**Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience.

# 4 Refining Visualizations with Sets and Parameters

**Lesson Time: 1 hour, 45 minutes**

## Lesson Introduction

It's important to organize data in specific ways to showcase insights and allows users to adjust aspects of visualizations to find the insights most relevant to them. In this lesson, you will create sets and analyze data using them in Tableau®. You will then apply parameters to refine visualizations.

## Lesson Objectives

In this lesson, you will:

- Create sets.
- Analyze data with sets.
- Apply parameters to data to refine visualizations.

# TOPIC A

## Create Sets

To discover the story data has to tell, it's often necessary to organize data in different ways. In this topic, you will create sets to organize data.

### Sets

***Sets*** organize data from the same or different dimensions in Tableau and can be used to filter what appears in the view. Sets are similar to groups, but are more functional. Where groups can organize data from a single dimension, sets can organize data from different dimensions.

For example, if you're analyzing product sales, you can choose to focus on analyzing only a subset of products, or combine products into categories for analysis.

You can create sets from the **Data** pane by right-clicking the dimension in the data pane that you wish to base the set on and selecting **Create→Set**, or by selecting ***marks*** in the Viz and selecting the **Create Set** icon from the tooltip.

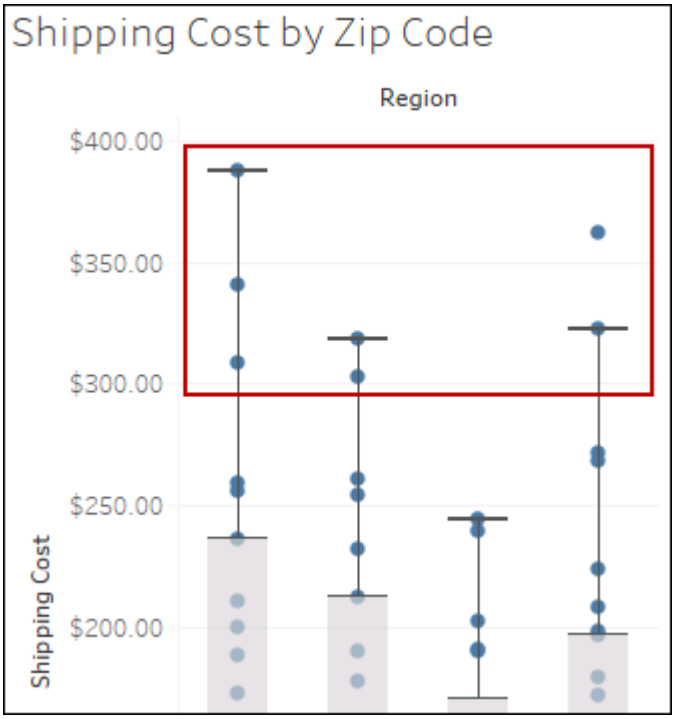



*Figure 4-1: Sets.*

You can create the following three types of sets.

| *Set Type* | *Description* |
|---|---|
| Constant or fixed sets | These are sets made up for marks selected in the Viz, or fields manually selected by the user. They are called constant or fixed because the members of the set do not change unless manually changed, even if the underlying data changes. |

| Set Type | Description |
|---|---|
| Dynamic sets | These sets have membership based on calculations. A conditional calculation is applied to the source dimension. To meet the requirements, members are included or excluded from the set. |
| Combined sets | Sets made of two sets based on the same dimension. This allows for different types of analysis, including cohort analysis. |

By default, sets include members based on the selections you make or calculation you create. You can specify that your set will exclude those members, as that may be easier in some cases.

## Calculated Members

***Calculated members*** are new fields you can create based on Tableau formulas or on ***Multidimensional Expressions (MDX)*** when using a multidimensional data source. There are two types of calculated members:

- ***Calculated measures*** are created based on an existing measure in the data with Tableau formulas or MDX applied to refine the data. For example, when looking at product sales, you might create a calculated member based on the sales measure to include only top-selling products.
- ***Calculated dimensions*** become a new member of an existing hierarchy. For example, the product sales data source might have a products dimension that contains all products sold and may include hierarchies for men's shoes and women's shoes. You may wish to create calculated members to break out dress shoes from athletic shoes. You can then perform an analysis of a calculated member by dragging it to the **Rows** shelf.

## Sets vs. Groups

Sets are similar to groups, so there are frequent questions about how sets differ from groups, and when each should be used.

Groups are used to combine field members from a single dimension into a category. For example, if you are looking at products in an office supply store, you might group desks, bookcases, and chairs into a group named Furniture. You may create other groups to organize office supplies like pens, staples, and paper, and electronics such as printer, computers, and copiers. The example also illustrates one of the primary uses for groups: to perform comparisons.

Sets, on the other hand, can be a little more difficult to create, but can provide much more functionality. Sets can organize members from single dimension or from multiple dimensions as part of a single set. For example, when analyzing sales, you can segment both products that have the highest sales volume and those that have the highest profit.

Sets can also be created from a calculation, making them dynamic so that membership changes when the underlying data changes. For example, you can create a calculated member to identify your most profitable products. The members of that set will change when the underlying data changes.

Sets, like groups, provide good comparisons for split or A/B testing, but they are also fundamentally different and powerful. Where groups combine items for comparison purposes, sets tend to be used for refining analysis by showing only the members or rows required for analysis. Sets can also be added to hierarchies. Also, while the member of a dimension can only be part of one category in a group, a member can belong to multiple categories in a set. For example, the member Notebooks could be a part of both School Supplies and Office Supplies. Sets can also be referenced by name in calculations, making it easy to use them in calculations.

## Set Icons

Sets are listed in the **Data** pane below measures. Different types of sets have different icons as listed in the following table.

| *Set Icon* | *Description* |
|---|---|
| | A set created by a Tableau user. |
| | Sets brought in to Tableau from a data source. |
| | A set created by Tableau automatically due to an action, such as a filter action. Users often don't need to interact with these sets as Tableau creates them to ensure filters are applied properly across views. |
| | A user defined filter of a set that is part of a published ***workbook***. |

## Constant or Fixed Sets

To create a constant or fixed set, select a **Dimension** in the **Data** pane, right-click and select **Create→Set**. On the **General** tab of the **Create Set** dialog box, you can opt to include all members of the dimension, select members from the dimension to include or exclude from the set, or create a custom list. When you're determining members of a set, you're telling Tableau what requirements are to be included in the set. Likewise, if you remove something from a set, you're removing it from the requirements for that set. To complete the creation of a set, you must also provide a name for the set.

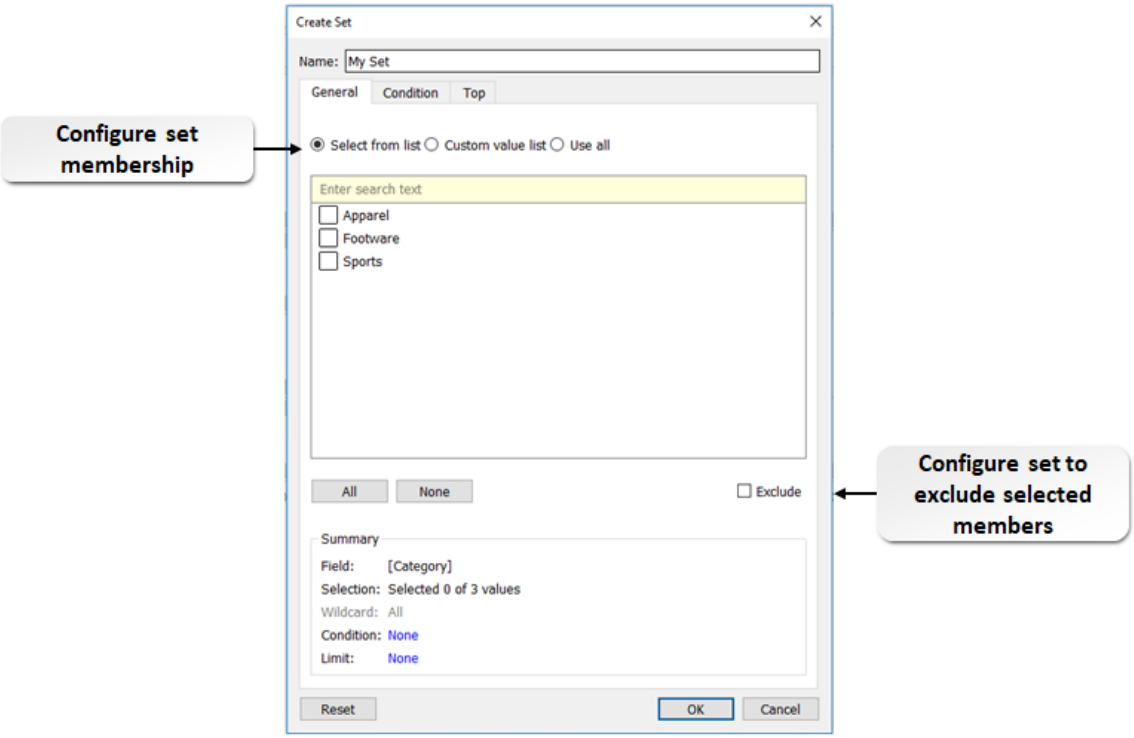



*Figure 4-2: Creating constant or fixed sets.*

Once you click **OK**, the set will appear under **Sets** in the **Data** pane. Any sets that you create will appear in the **Data** pane, in the **Sets** section below **Measures**.

## Sets From Marks

You can create sets from marks in the Viz by selecting the marks with the mouse, right-clicking and selecting the **Create Set** icon. Grabbing marks from the Viz creates a constant or fixed set. It is an easy way to include members from multiple dimensions and measures. For example, if you're doing a magic quadrant analysis, you might show the best-selling and highest-profit products in the upper-right of the Viz. You can grab all the marks in that quadrant to do further analysis and find insight about why those products sell well or are highly profitable.

When you create, a set you must give it a name at a minimum, but you can also configure other options. You can select other marks in the Viz after you've created a set. If you do, Tableau gives you the option to include or exclude those marks from the sets already created or to create a new set.

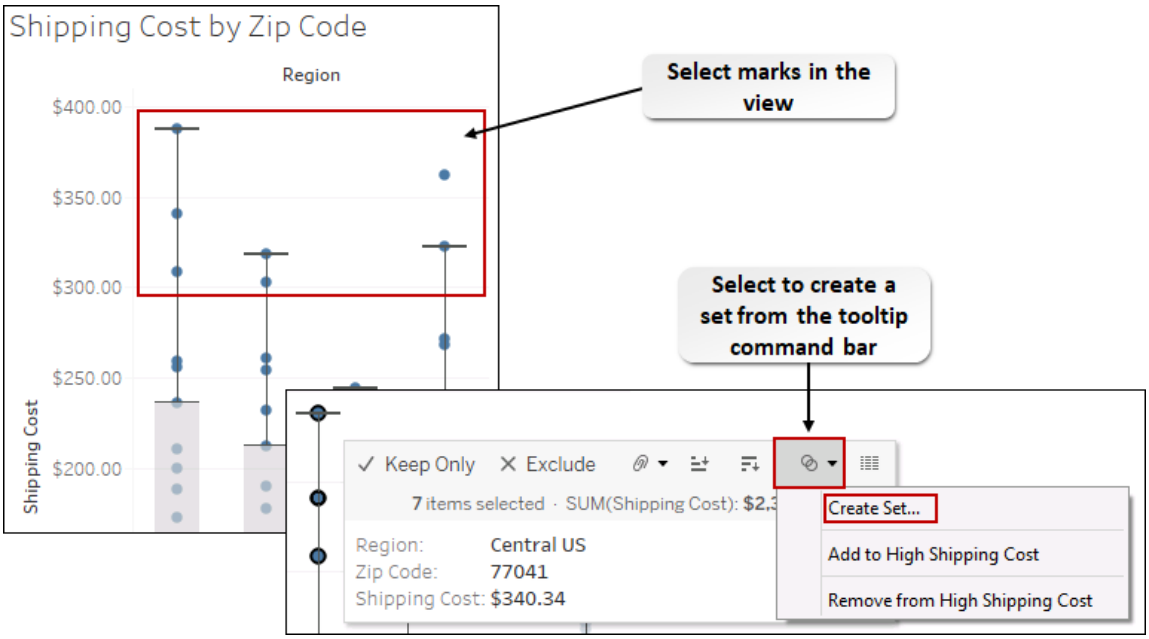



*Figure 4-3: Sets from marks.*

## Calculated or Dynamic Sets

To create a calculated set, start by selecting a dimension in the **Data** pane and creating the set the same way you would create a fixed set. This time, however, you will us the **Condition** tab to configure which rules determine which members of the dimension the set will include or exclude. You can add rules by selecting the field and arithmetic operators to determine the members of the set. Specify a range of values or enter a formula.

For example, if you wanted to see sales over $10,000, you could create a rule using sum of the **Sales** field for any items greater than or equal to $10,000. You can also enter a formula using Tableau formulas and box dot notation.

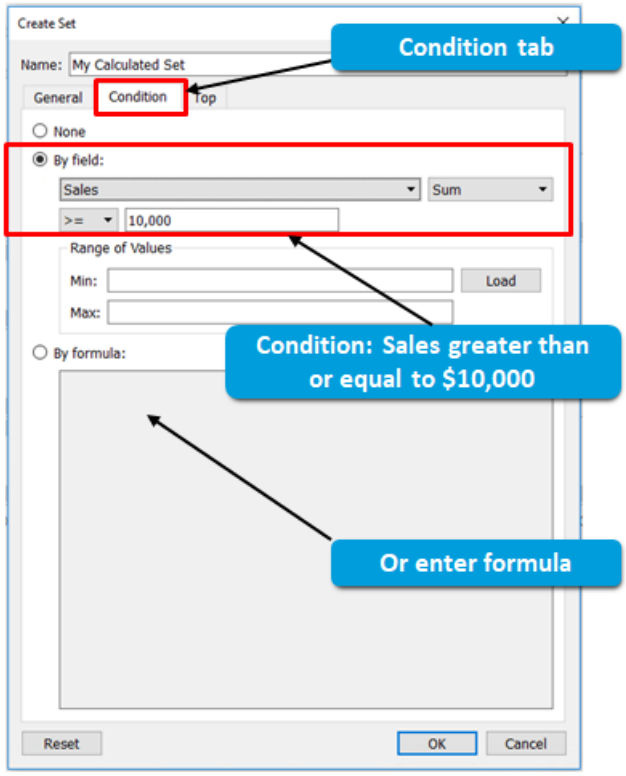



*Figure 4-4: A set calculated by condition.*

You can also select the **Top** tab in the **Create Set** dialog box to configure membership based on the top values in the field on which the calculated set is based.

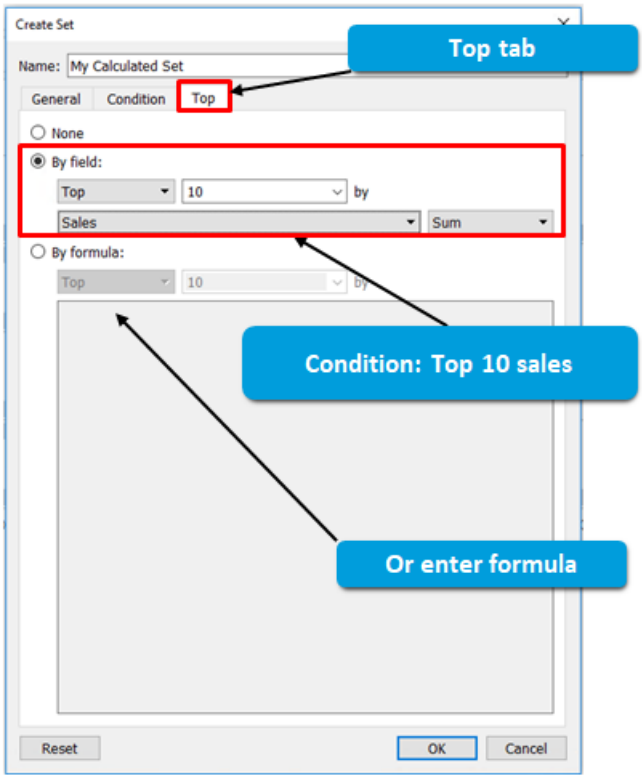



*Figure 4-5: A set calculated by top members.*

## Combined Sets

You can only combine sets based on the same dimension. For example, if you wanted to look at products, you might have a set to see which products have the lowest shipping costs and highest number of sales. You can create one set to show products with shipping costs below one dollar, and another to show products with sales over $100,000. Since both sets are based on the product sales dimension, you can combine the sets.

When you combine sets, you are effectively using a join to determine which members should be a part of the set. You can select:

- **All members in Both Sets**: This is similar to an outer join.
- **Shared members in Both Sets**: This is similar to an inner join.
- **<Left Set Name> except shared members**: Include all members in the set identified on the left side of the dialog box, excluding any members it shares with the set named on the right side of the dialog box.
- **<Right Set Name> except shared members**: Include all members in the set identified on the right side of the dialog box, excluding any members it shares with the set named on the left side of the dialog box.

## In/Out vs. Members

Sets make it easy to work with their members in the view. For example, if you drag a set to **Color** on the **Marks** *card*, the set is placed in **In/Out mode**, and members of the set are coded to one color while all other marks are coded another color. You can manually turn **In/Out mode** on by right-clicking the set and selecting **In/Out mode**.

You can also select to **Show Members in Set** from the right-click menu. This is similar to applying a filter as only members of the set will be displayed in the view.

These options allow you to quickly identify the data you're most interested in based on the sets you've created.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Create and Combine Sets.** |

# ACTIVITY 4-1
## Creating and Combining Sets

### Data File
C:\095205Data\Refining Visualizations with Sets and Parameters\MyFootPrintSports_2018_4.twb

### Before You Begin
Tableau Desktop is open.

### Scenario
As you have used Tableau, you have found there are times when you want to isolate data so you can further investigate the associated numbers. There have been times when creating filters or manually configuring marks to fit your needs have been a big effort. You want to explore sets and see if they can aid you in isolating data and then apply them to different views. You want to examine your shipping costs and isolate areas that have had shipping costs. You want to look at the total cost and the average cost.

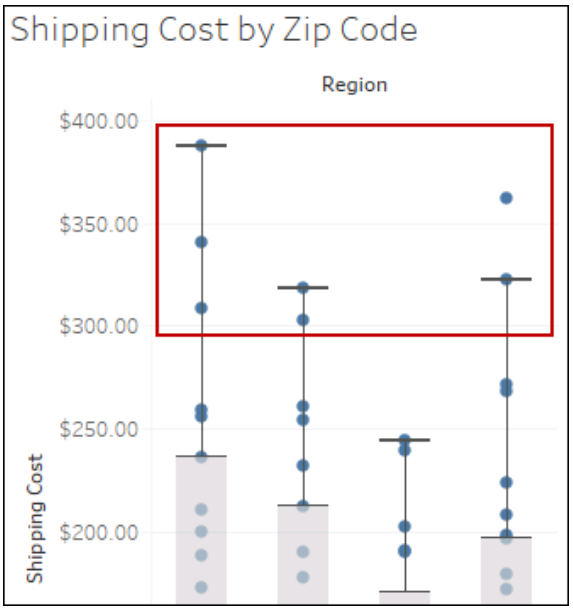
---

1. Open the worksheet for the lesson.
   a) From the menu, select **File→Open**.
   b) In the **Open** dialog box, browse to the **C:\095205Data\Refining Visualizations with Sets and Parameters** folder.
   c) Select **MyFootPrintSports_2018_4** and select **Open**.

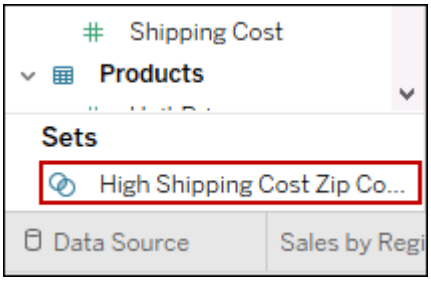
2. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Refining Visualizations with Sets and Parameters** folder.
   c) In the **File name** box, type *MyCopy_MyFootPrintSports_2018_4* and select **Save**.

3. Create a set from marks.
   a) At the bottom of the screen, select the **Shipping Cost by Zip Code** worksheet.

b) In the canvas, select the marks above the **$300.00 Shipping Cost** line by drawing a rectangle with the mouse.




c) Right-click one of the selected marks and select **Create Set**.
d) In the **Create Set** dialog box, observe the regions and ZIP Codes included in this set.
e) In the **Name** box, type ***High Shipping Cost Zip Codes***
f) Select **OK**.
g) At the bottom of the **Data** pane, under **Sets**, observe the newly created **High Shipping Cost Zip Codes** set.



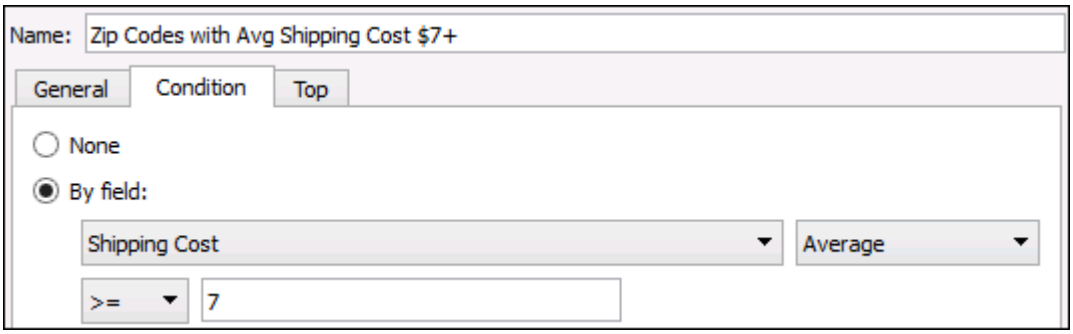

## 4. Use the **High Shipping Cost Zip Codes** set in a view.

a) At the bottom of the screen, select the **Average Shipping Cost and Profit** worksheet.
b) In the **Data** pane, under **Sets**, select and drag **High Shipping Cost Zip Codes** to the **Color** box in the **Marks** card.
c) Select the color box in the **Marks** card and select **Edit Colors**.
d) In the **Edit Colors [In/Out of High Shipping Cost Zip Codes]** dialog box, under **Select Data Item**, select **Out**.
e) Under **Select Color Palette**, select the **Red** square or some other bright color to contrast with the **In** color.
f) Select **OK**.

g) In the canvas, observe the change to the scatter plot. Those seven ZIP Codes are now a different color and are in the **In** set. The other marks are in the **Out** set. While those ZIP Codes had the largest sums of total shipping costs, many orders in those ZIP Codes do not have the highest average shipping cost.

5. Create a calculated set.

a) In the **Data** pane, under **Dimensions→Location**, right-click **Zip Code** and select **Create→Set**.
b) In the **Create Set** dialog box, in the **Name** box, type ***Zip Codes with Avg Shipping Cost $7+***
c) Select the **Condition** tab.
d) Select **By field**.
e) In the **Category** drop-down list, select **Shipping Cost**.
f) In the **Sum** drop-down list, select **Average**.
g) In the **=** drop-down list, select **>=**.
h) In the box, type ***7***




i) Select **OK**.

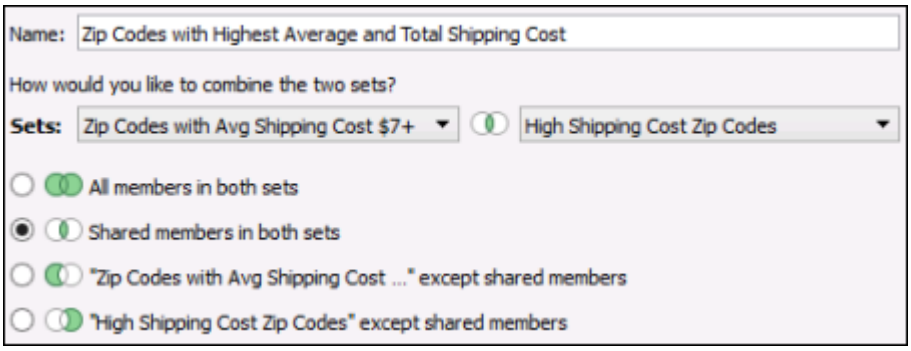
6. Use the **Zip Codes with Avg Shipping Cost $7+** set in a view.

a) At the bottom of the screen, verify the **Average Shipping Cost and Profit** tab is selected.
b) In the **Data** pane, under **Sets**, select and drag **Zip Codes with Avg Shipping Cost $7+** to the **Color** box in the **Marks** card. This replaces the previous set that was in the **Color** box.
c) Select the color box in the **Marks** card and select **Edit Colors**.
d) In the **Edit Colors [In/Out of High Shipping Cost Zip Codes]** dialog box, under **Select Data Item**, select **Out**.
e) Under **Select Color Palette**, select the **Red** square or some other bright color to contrast with the **In** color.
f) Select **OK**.
g) In the canvas, observe the change to the scatter plot. ZIP Codes with an average shipping cost of $7 or more are now a different color.

7. Combine two sets.

a) In the **Data** pane, under **Sets**, right-click **Zip Codes with Avg Shipping Cost $7+**, and then select **Create Combined Set**.
b) In the **Name** box, type ***Zip Codes with Highest Average and Total Shipping Cost***
c) In the **Sets** drop-down list on the right, select **High Shipping Cost Zip Codes**.

d) Select **Shared members in both sets**.




e) Select **OK**.

8. Use the combined **Zip Codes with Highest Average and Total Shipping Cost** set in a view.

a) At the bottom of the screen, verify the **Average Shipping Cost and Profit** tab is selected.

b) In the **Data** pane, under **Sets**, select and drag **Zip Codes with Highest Average and Total Shipping Cost** to the **Color** box in the **Marks** card. This replaces the previous set that was in the **Color** box.

c) Select the color box in the **Marks** card and select **Edit Colors**.

d) In the **Edit Colors [In/Out of High Shipping Cost Zip Codes]** dialog box, under **Select Data Item**, select **Out**.

e) Under **Select Color Palette**, select the **Red** square or some other bright color to contrast with the **In** color.

f) Select **OK**.

g) In the canvas, observe the change to the scatter plot. ZIP Codes with an average shipping cost of $7 or more and ZIP Codes identified as one of the seven with the highest shipping costs are now a different color.

9. Show only members of a set.
   a) In the **Marks** card, right-click the **IN/OUT(Zip Codes with Highest Average and Total Shipping Cost)** pill, and then select **Show Members in Set**.
   b) In the canvas, observe that only the three ZIP Codes from the set are now displayed.

# TOPIC B

## Analyze Data with Sets

Once you've organized data by creating sets, Tableau provides many ways you can use sets as part of your data analysis. In this topic, you will analyze data using sets.

### Options for Editing Sets

Once created, you can edit a set by right-clicking the set in the **Data** pane and selecting **Edit Set**.

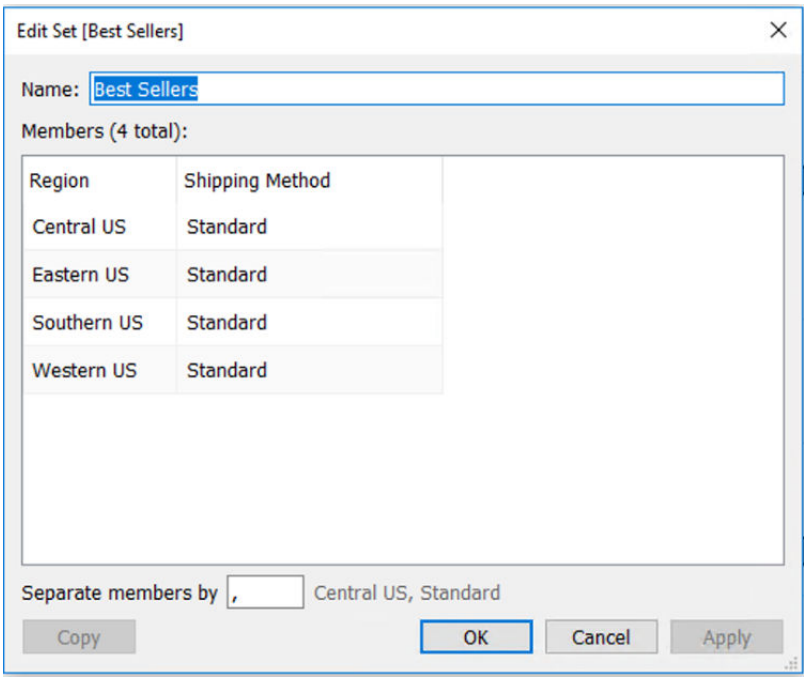



*Figure 4-6: Editing a manually created set.*

The degree to which the set can be edited, and the modifications you can make while editing, depends on the type of set:

- **Fixed set**: If a set was created from marks or is a manually created fixed set, you can only see the members of the set in the **Edit Set** dialog box. You can remove members of the set. You cannot add new members. If requirements for the set have changed, you can potentially create a new set based on the same dimension, add new members, and combine the new set with the old set.
- **Dynamic set**: With dynamic sets, you have all the options available when you edit the set as you have when you create the set. You can change the field calculation or adjust the formula, and the set membership will be adjusted to reflect the changes once editing is complete.
- **Combined set**: Combined sets, like dynamic sets, have all the options available when you edit the set as you have when you create the set. You can reconfigure which sets are being combined in the set, and adjust the relationship of the members between the sets that will make up the combined set.

### Create Set from Filtered Results

A common reason for creating sets is to replace commonly used filter results. For example, if a view has multiple filters that can be configured, and you frequently configure those filters to show a

specific subset of marks, it may be worthwhile to create a set based on that filter configuration. You can do this either by creating a dynamic set, or by configuring the filters in the view and creating a set by selecting the visible marks.

For example, if you're analyzing how product profitability in North America is related to shipping costs, you might configure the filters for the view to show results for the United States and Canada for products with a profit of one cent or more, and shipping costs below a dollar. Creating a set to reflect these filter settings would save time setting up the filters.

You can also create a set from a single, specific filter by right-clicking the pill on the filter shelf and selecting **Create Set**.

## Sets in Calculations

Sets can also be named directly in calculations. This makes it easy to make a calculated field using data from sets. For example, if you created a set to show profitable products in North America, with shipping costs below one dollar, and named it North America Profitable Products with Low Shipping, you could create a calculated field that references that set directly to subtract the shipping costs and just see the profit:

```
SUM[North America Profitable Products with Low Shipping] - SUM[Products].
[ShippingExpenses]
```

These types of calculations can be used in many ways. If you create sets to segment based on location, customer name, or product type, you can use those sets to create calculated fields that give you insights about the sets.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Analyze Data with Sets.** |

# ACTIVITY 4-2

## Analyzing Data with Sets

### Before You Begin

The **MyCopy_MyFootPrintSports_2018_4** workbook is open in Tableau Desktop.

### Scenario

You like the results of your first attempt with sets, but you want to make some changes. You want to edit the set to include average shipping costs of $6 and higher. Then you want to see if you can duplicate the effects of a filter using a set. This filter will only include the shipping costs for the standard and two-day shipping methods, and regions that had $800 or more in shipping costs. Then you can use that set as a consistent filter you can apply to future views.

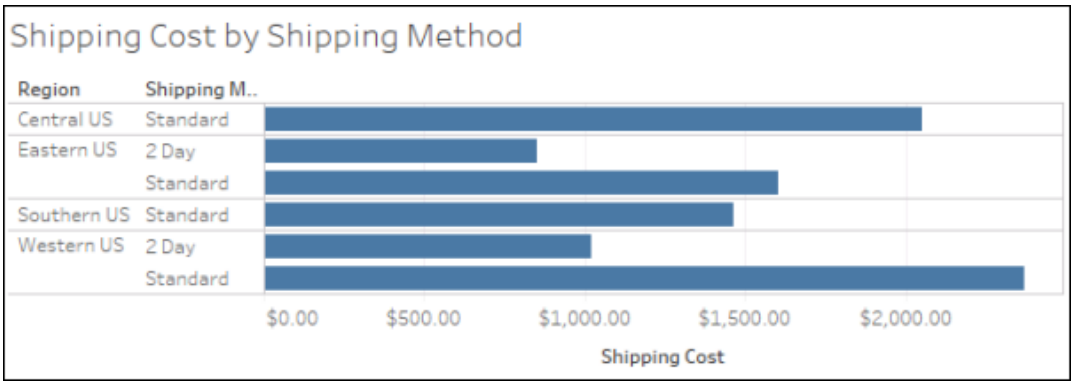
---

1. Try to edit a combined set to change the average shipping cost setting.
   a) In the **Data** pane, under **Sets**, right-click **Zip Codes with Highest Average and Total Shipping Cost**, and then select **Edit Set**.
   b) Observe the settings you can change. In the combined set, you can only change the settings for combining the sets. To edit the settings for the sets, you need to edit the original set. Select **Cancel**.

2. Edit a set.
   a) In the **Data** pane, under **Sets**, right-click **Zip Codes with Avg Shipping Cost $7+**, and then select **Edit Set**.
   b) In the **Edit Set [Zip Codes with Avg Shipping Cost $7+]** dialog box, in the **7** box, type *6*
   c) In the **Name** box, change the name to *Zip Codes with Avg Shipping Cost $6+*
   d) Select **OK**.
   e) In the canvas, observe there are now seven ZIP Codes listed in which the average shipping cost was lowered to $6.

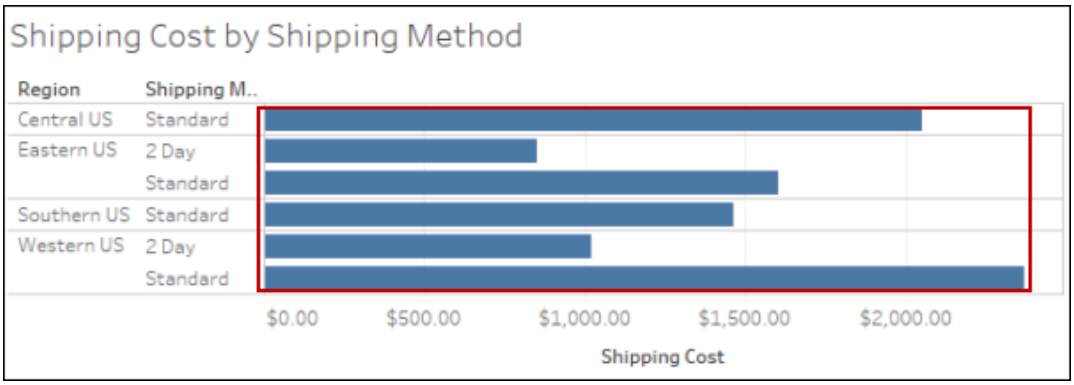
## 3. Add filters to a view.

a) At the bottom of the screen, select the **Shipping Costs by Shipping Method** worksheet.
b) In the **Data** pane, under **Dimensions→Orders**, select and drag **Shipping Method** to the **Filters** card.
c) In the **Filter [Shipping Method]** dialog box, in the list, uncheck the **First Class** and **Next Day** check boxes.
d) Select **OK**.
e) In the **Data** pane, under **Measures→Orders**, select and drag **Shipping Cost** to the **Filters** card.
f) In the **Filter field [Shipping Cost]** dialog box, in the list, select **Sum**, and then select **Next**.
g) In the **Filter [Shipping Cost]** dialog box, select **At least**.
h) In the box, type ***800***. Note this shipping cost is per shipping method for a region, not a ZIP Code as in the previous activity.
i) Select **OK**.
j) In the canvas, observe the changes to the results.




## 4. Create a set from a filter.

a) In the canvas, select all of the marks in the view by drawing a rectangle with the mouse.



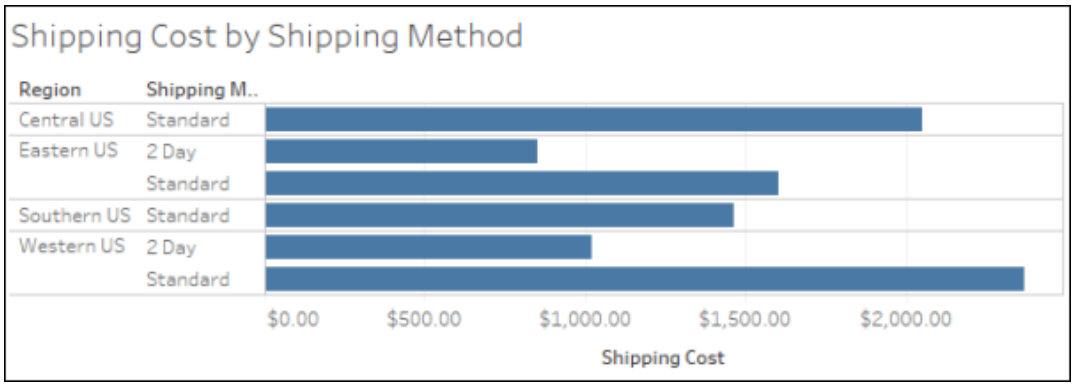

| | |
|---|---|
| | **Note:** Start drawing the rectangle to the right of either the first or last bar mark. |

b) Right-click one of the selected marks and select **Create Set**.
c) In the **Create Set** dialog box, in the **Name** box, type ***Shipping Method Filters***
d) Check the **Add to filters shelf** check box.
e) Select **OK**.

## 5. Test a set created from a filter.

a) In the **Filters** card, right-click **Shipping Method**, and then select **Remove**.
b) In the **Filters** card, right-click **SUM(Shipping Cost)**, and then select **Remove**.

c) In the canvas, observe that the results are unchanged and the set is duplicating the filters that were in place previously.



# TOPIC C

## Apply Parameters to Data to Refine Visualizations

Once you have your data organized and are using it as part of your analysis, it can be helpful to allow people who are viewing your visualizations to adjust how that data is displayed so they can showcase information that is more important to them. In this topic, you will apply parameters to data to refine visualizations.

### Parameters

A ***parameter*** defines a variable similar to those in equations, that can be used to control how a defined element is displayed in the view. Parameters can be added to a workbook to add interactivity and flexibility to visualizations. Parameters are analogous to a variable in an equation. With parameters, the user controls the value of that variable. Unlike filters, which remove data from the view, parameters adjust the output of another defined element, such as marks that are colored or the value of a ***reference line***. For example, parameters can be used to color marks in the view. By selecting the value of the parameter, a user can define a broader or narrower set of color marks. In another example, a user can set a reference line for the view based on the parameter's value.

Parameter controls, like filters, are placed at the right side of the Viz where users can interact with them. It's important to conceptually understand that parameters by themselves don't impact the view. Since they control a variable value, they can only impact the view once that value is associated with something else like a filter, set, reference line, or calculated field. In fact, if a parameter is created and associated with something in the Viz, and that element is later removed, Tableau will label that parameter as an orphaned parameter in the Viz.

Once defined, parameters can be selected and edited in the **Data** pane below **Sets**. Parameters are available workbook-wide and can be used in multiple places.

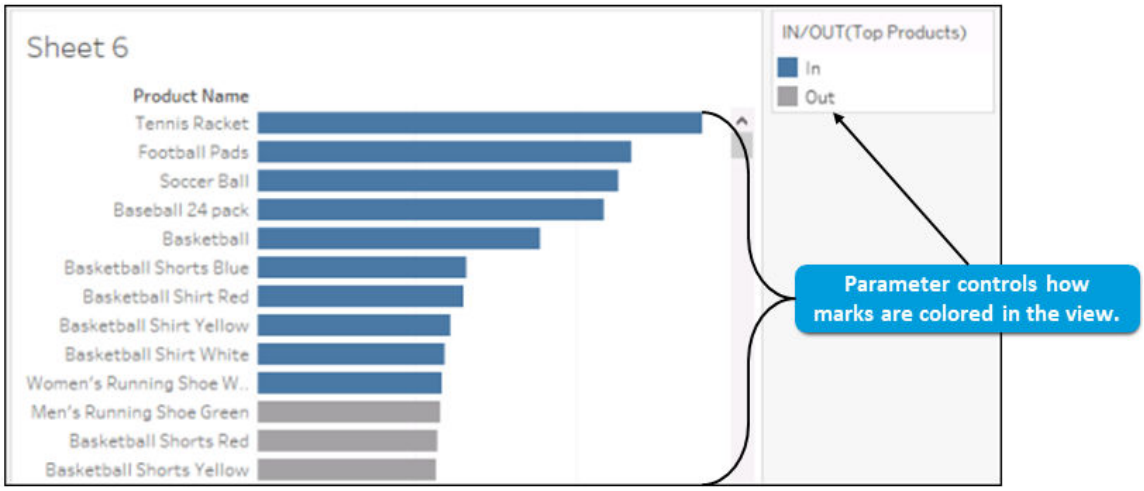



*Figure 4-7: Parameters.*

### Parameter Types

The type of parameter you create affects where it can be used and the display formats available.

| Type | Description |
|---|---|
| **Float** | Returns a decimal number, such as 1.5. |
| **Integer** | Returns an integer number with no decimals. For example, 1.5 is rounded to 2. |
| **String** | Returns text, such as "Profitable." |
| **Boolean** | Returns true or false. |
| **Date** | Returns a date number, string, or data expression, such as 8/27/2018. |
| **Data & Time** | Returns a data and time number, string, or data and time expression, such as 8/27/2018 10:16:00. |

## Create Parameters

To create a parameter from scratch, in the **Data** pane, right-click and select **Create Parameter**. You must define the following in the create parameter dialog box.

| Property | Description |
|---|---|
| **Name** | Name the parameter something meaningful. |
| **Data type** | Select the type of parameter you wish to create based on what the parameter will do, such as floating or Boolean. |
| **Current value** | This is the default value for the parameter. |
| **Display format** | Format values based on the type of parameter you're creating. This will affect what it will look like in the Viz including decimal places, negative values, day-date display, and currency configuration. |
| **Allowable values** | You can select from all values, a list, or a range where you can configure minimum and maximum values, as well as a step size for the slider in the Viz. |
| **Set from parameter** | Allows you to configure this parameter to be set based on how a different parameter is configured. |
| **Set from field** | Allows you to configure this parameter to be set based on the value of a field. |

## Parameters with Filters

You can use parameters with filters by right-clicking the filter and selecting **Edit Filter**. For example, you may have to create a filter to display the top-performing stores based on gross sales. However, the filter may be configured on the **Top** tab to display the 10 top performers. In the **Top** tab of the filter, rather than configure the number of top performers to show, you can create a new parameter to display a range of stores based on their gross sales from greatest to least. You can also see and select any other parameters defined in the workbook. Once configured, the parameter control will display automatically in the view.

Basing this value on a parameter allows each user to adjust how many top performers they wish to see. So, if a user is only interested in the top five performers, they can adjust the parameter accordingly. A different user might wish to see the top 20 performers.

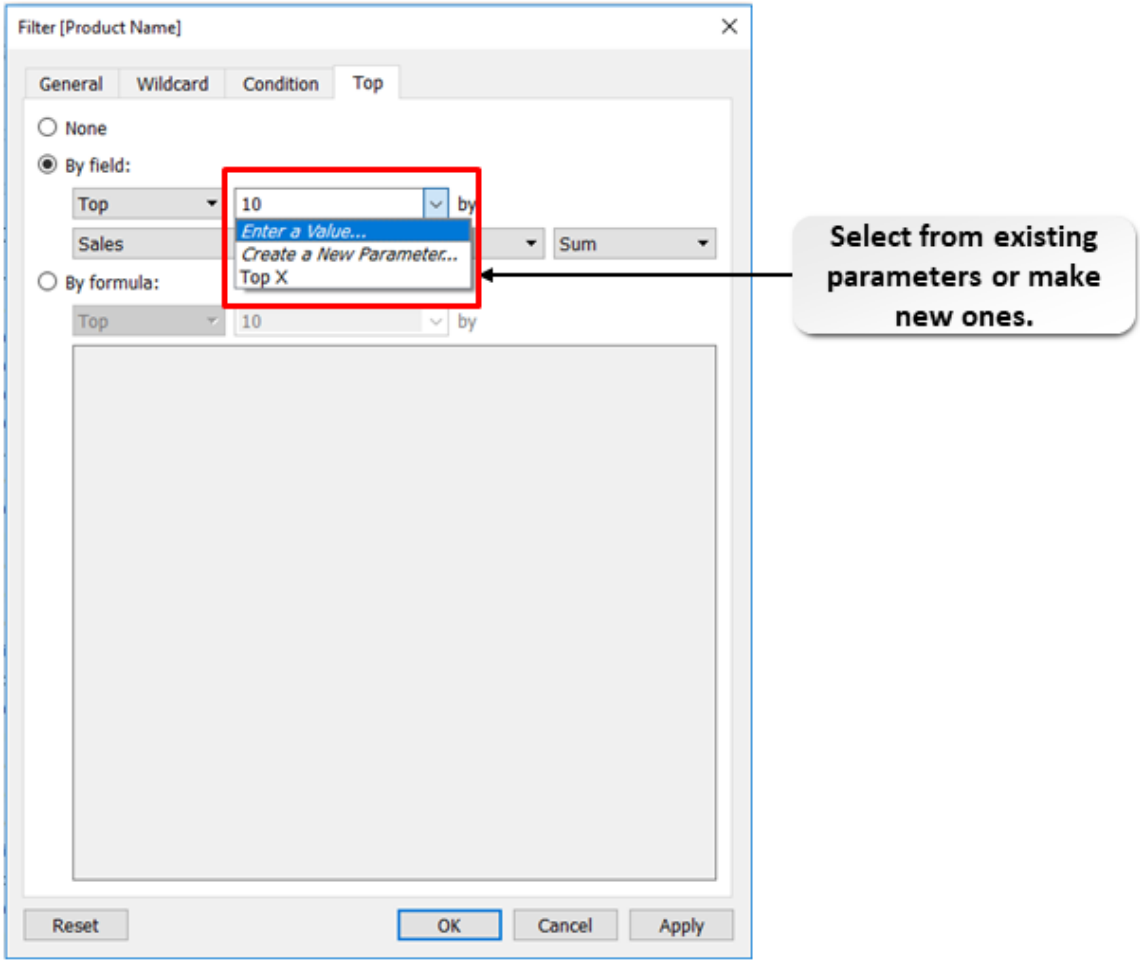



*Figure 4-8: Parameters with filters.*

## Parameters with Sets

Similarly, you can use parameters to define membership in a set. Select or create a parameter in the **Top** tab or the **Create Set or Edit Set** dialog box. In this case, the parameter control may not display automatically in the view, but you can right-click the control and select **Show Parameter Control**. Users can then adjust the parameter to define the membership in the set that will affect the marks displayed in the view.

## Parameters with Reference Lines

You can use parameters with reference lines to give users the ability to adjust the value of the reference line in relation to marks in the view based on the analysis being performed. Once you add a reference line from the **Analytics** tab, instead of configuring a specific value, you can create or select a parameter. For example, if you wanted to analyze product profitability based on shipping costs, you could create a reference line based on a parameter that allows you to adjust the shipping cost. By adjusting the parameter, you can easily see what products are profitable at higher and lower shipping costs.

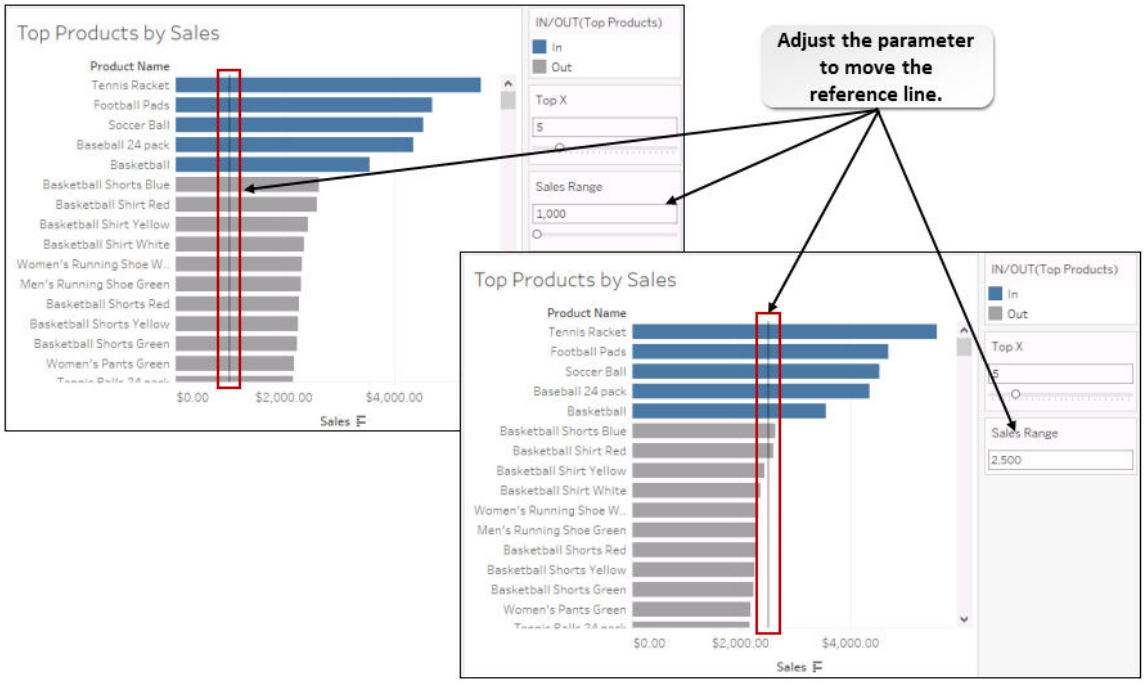



*Figure 4-9: Parameters with reference lines.*

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Apply Parameters to Data to Refine Visualizations.** |

# ACTIVITY 4-3

## Applying Parameters to Data to Refine Visualizations

### Before You Begin

The **MyCopy_MyFootPrintSports_2018_4** workbook is open in Tableau Desktop.

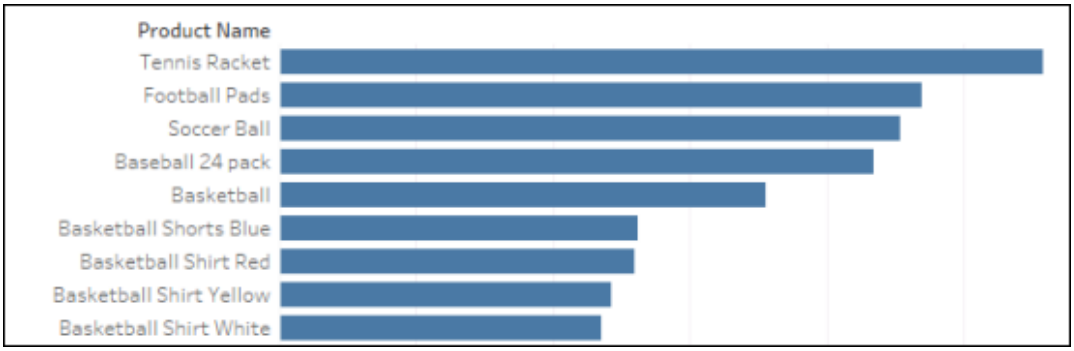
### Scenario

You want to be able to view the top products based on sales. You also want to be able to easily change the number of top products included in the view. You will use a parameter and a set to accomplish this. You also want to mark a threshold in product sales using a reference line. You are not sure of the exact sales amount you want for the reference line, so you will use a parameter so you can adjust it until you find the threshold you want.

---

1. Create a worksheet for top products.

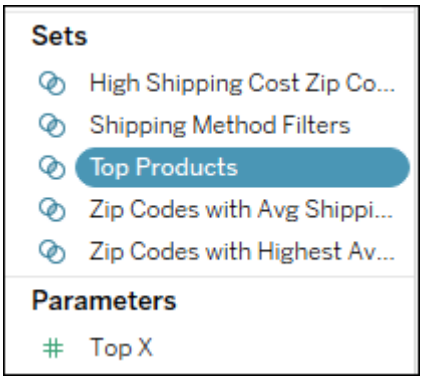
   a) On the toolbar, select the **New Worksheet** button.
   b) On the **Sheet 6** worksheet tab, in the **Data** pane, under **Dimensions→Products**, select and drag **Product Name** to the **Rows** shelf.
   c) In the **Data** pane, under **Measures→Orders**, select and drag **Sales** to the **Columns** shelf.
   d) Hover the mouse pointer over the **Sales** axis and select the **Quick Sort** button to sort the product list in descending order.
   e) Observe the list of products sorted by sales.



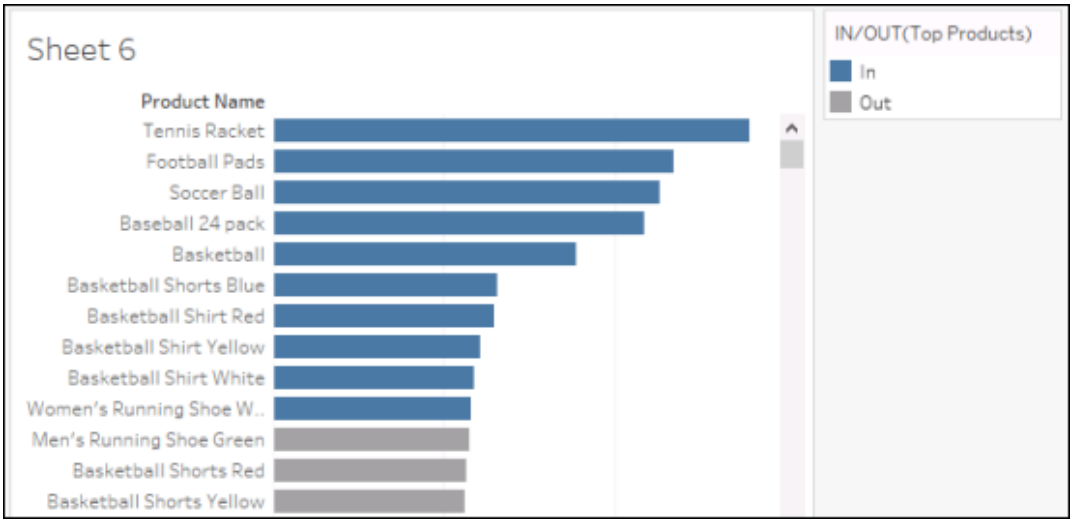

2. Create a parameter and a set for top products.

   a) In the **Data** pane, under **Dimensions→Products**, right-click **Product Name** and select **Create→Set**.
   b) In the **Create Set** dialog box, in the **Name** box, type *Top Products*
   c) Select the **Top** tab.
   d) Select **By field**.
   e) In the **10** drop-down list, select **Create a New Parameter**.
   f) In the **Create Parameter** dialog box, in the **Name** box, type *Top X*
   g) In the **Range of values** section, for **Maximum**, type *25*
   h) Select **OK** twice.

i) In the **Data** pane, verify **Top Products** appears under **Sets**, and **Top X** appears under **Parameters**.

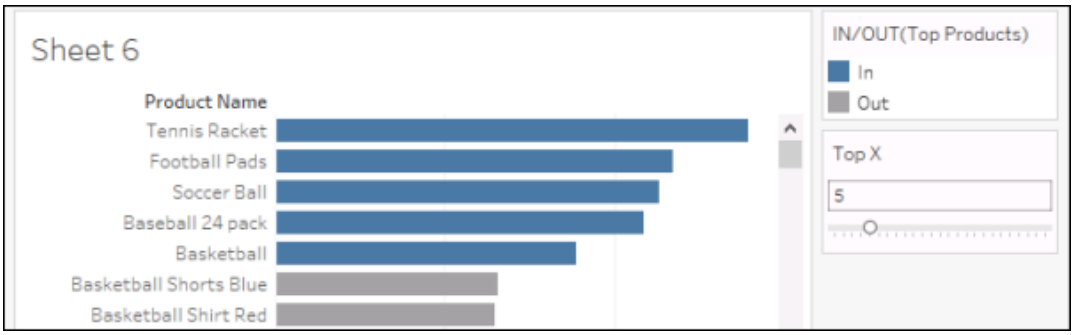



3. Add the **Top Products** set to the view.
   a) In the **Data** pane, under **Sets**, select and drag **Top Products** to the **Color** box in the **Marks** card.
   b) Observe the list of products and that the top 10 have been highlighted.



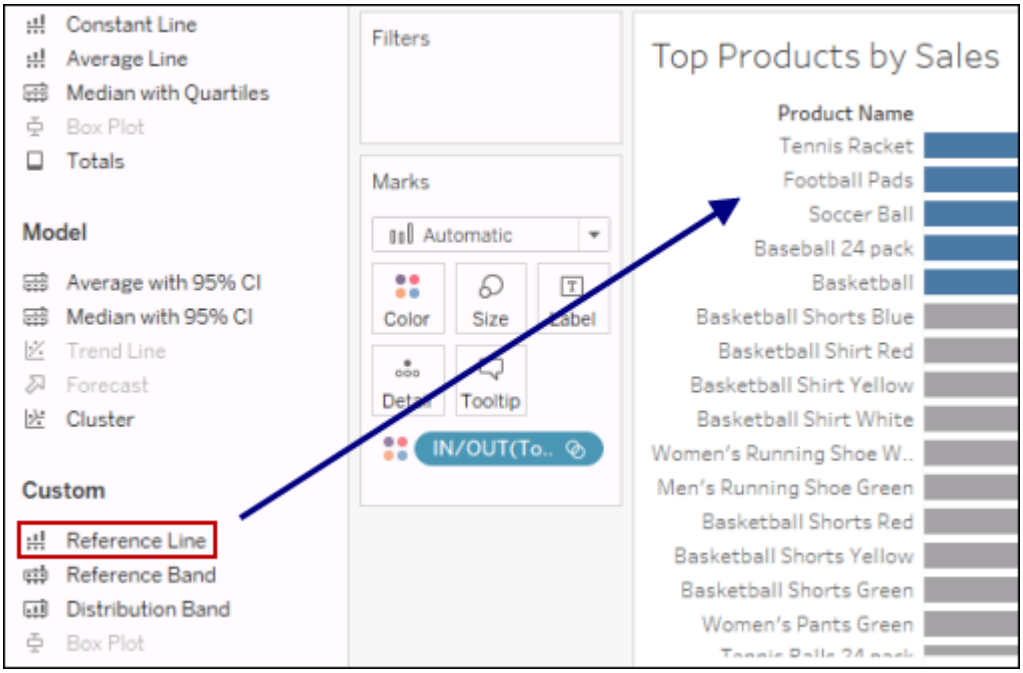

4. Use the top X parameter.
   a) In the **Data** pane, under **Parameters**, right-click **Top X** and select **Show Parameter Control**.
   b) In the legend, in the **Top X** card, use the slider to change the value to **25**.
   c) Observe that the top 25 products have been highlighted.
   d) In the legend, in the **Top X** card, use the slider to change the value to **5**.

e) Observe that the top five products have been highlighted.



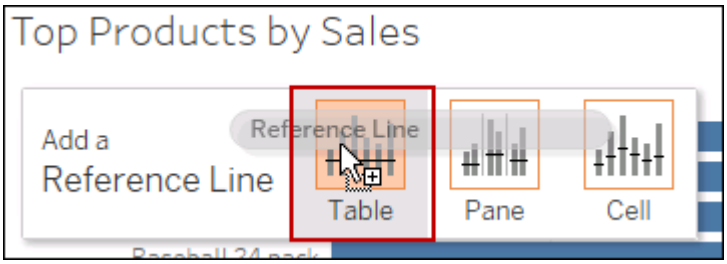

5. Rename the worksheet.
   a) Right-click the **Sheet 6** tab and select **Rename**.
   b) Type ***Top Products by Sales*** and press **Enter**.

6. Add a reference line that uses a parameter.
   a) At the top of the **Data** pane, select the **Analytics** tab.
   b) In the **Analytics** pane, under **Custom**, select and drag **Reference Line** to the canvas and hover the mouse pointer over it without dropping it.




| | |
|---|---|
| | **Note:** When dragging analytics items to the canvas, you need to hover the mouse pointer over the canvas and wait for the menu to appear. |

c) From the menu, drop the **Reference Line** on **Table**.

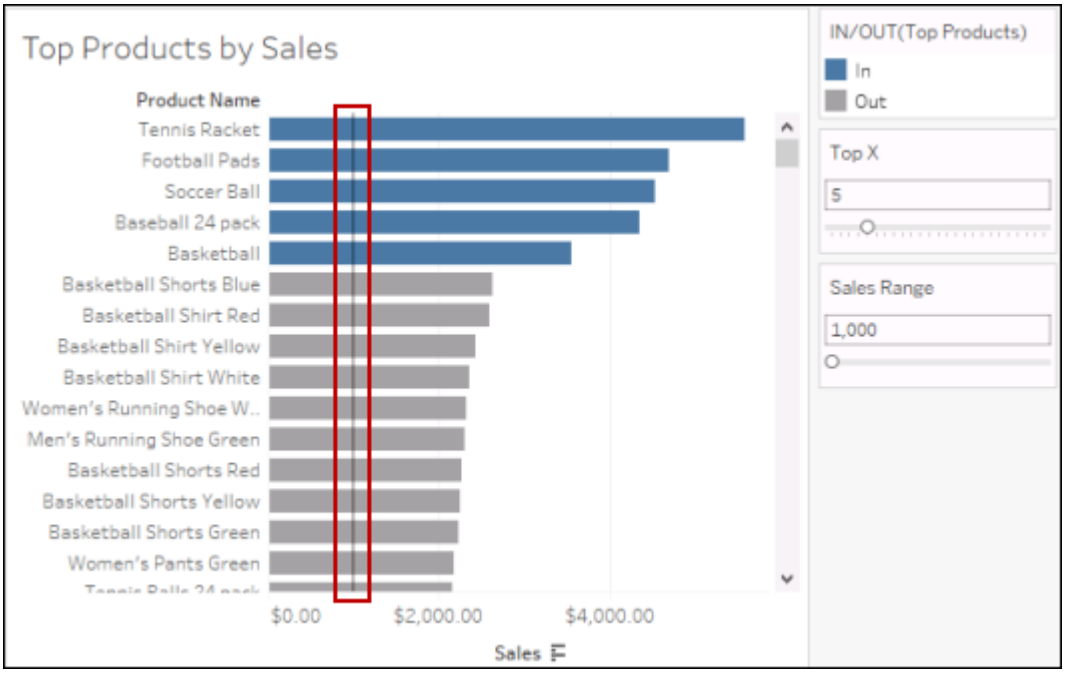



> **Note:** If you dropped the reference line before dropping it on **Table**, then continue with the steps as written. The next step verifies you have the correct scope selected.

d) In the **Edit Reference Line, Band, or Box** dialog box, for **Scope**, verify **Entire Table** is selected.
e) In the **Line** section, in the **Value** drop-down list, select **Create a New Parameter**.
f) In the **Create Parameter** dialog box, in the **Name** box, type ***Sales Range***
g) For **Allowable values**, select **Range**.
h) In the **Range of values** section, check the **Minimum** check box.
i) In the **Minimum** box, type ***1000***
j) Check the **Maximum** check box.
k) In the **Maximum** box, type ***4000***
l) Select **OK** twice.

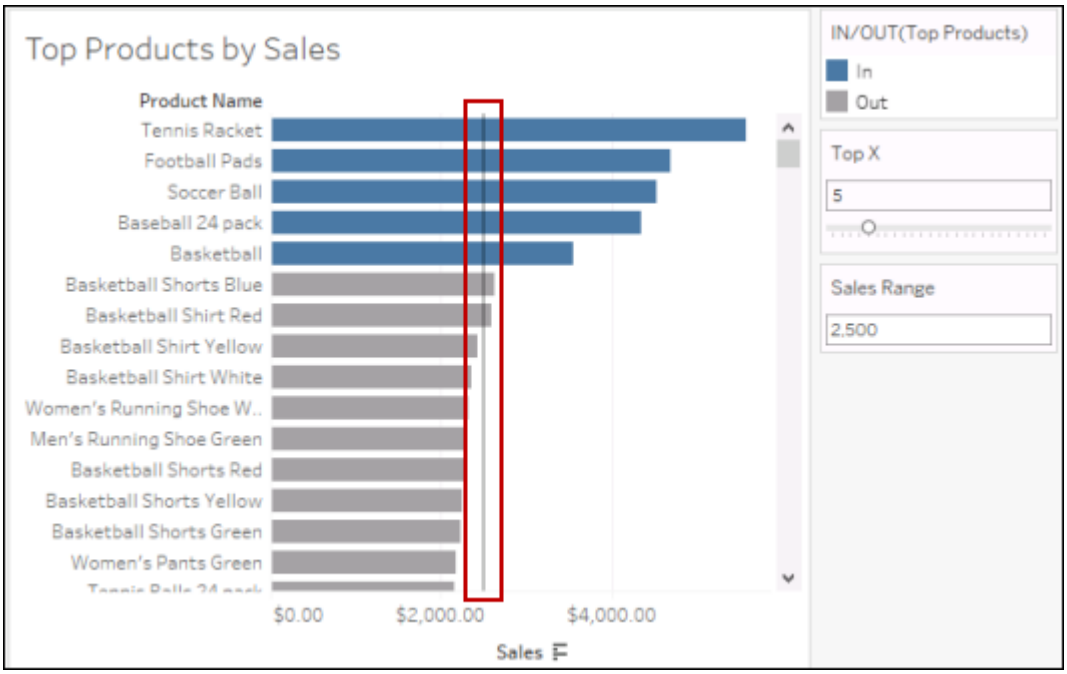
## 7. Use the reference line.

a) Observe the reference line is set at $1,000.




b) In the legend, in the **Sales Range** card, select the drop-down arrow and select **Type In**.
c) In the legend, in the **Sales Range** card, change the value to ***2500***

d) Observe the reference line is now set at the desired sales amount of $2,500.




e) In the legend, in the **Top X** card, use the slider to change the value to **7** to match the cut off of the reference line.

## 8. Save the workbook.

a) From the menu, select **File→Save**.

b) From the menu, select **File→Close**.

# Summary

In this lesson, you created fixed and dynamic sets to organize data for analysis. You used sets in calculations and to filter results in the view. You also applied parameters to adjust data in the view.

### In your environment, how might you use sets to organize your data?

### How might your organization use parameters in Tableau?

| | |
|---|---|
| | **Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience. |

# 5 Manipulating Data with Calculations

**Lesson Time: 2 hours**

## Lesson Introduction

Connecting to and extracting from data sources is often only the first step in data analysis or creating visualizations. It's often necessary to manipulate that data to show insights and relationships. In this lesson, you will create calculated fields to find insights, use functions to manipulate data, and analyze data with table calculations in Tableau®.

## Lesson Objectives

In this lesson, you will:

- Create calculated fields.
- Manipulate data with calculations.
- Analyze data with table calculations.

# TOPIC A

## Create Calculated Fields

To gain new insights from data brought into Tableau, it's often necessary to perform calculations on that data. In this topic, you will create calculated fields.

### Calculated Fields and Calculations

In many instances, the data source you are working with either won't have all the data you need to perform an analysis or show the insights you wish, or the data that exists will need to be refined. In those cases, you can create ***calculated fields*** to obtain the data you need to perform your analysis.

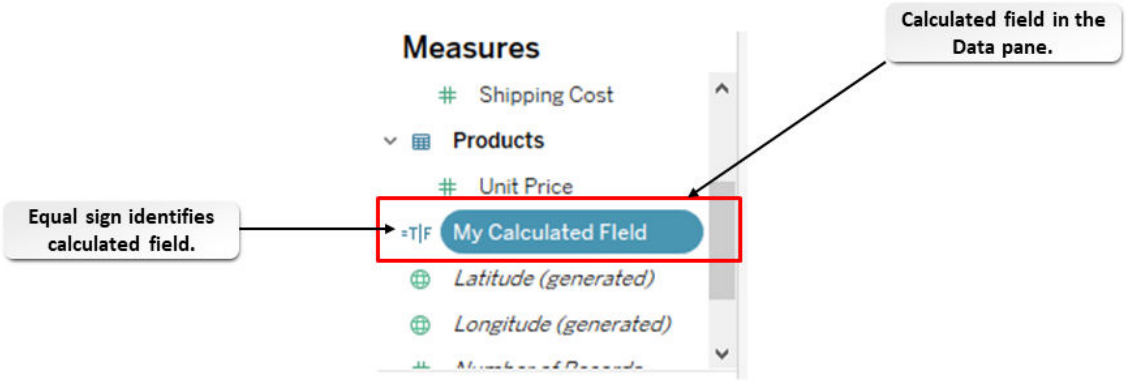
As the name implies, a calculated field is created by building formulas using data from fields, basic arithmetic operators such as the plus ( + ) and minus ( – ) symbols, and using functions built into Tableau. The formulas used to create a calculated field is similar to formulas used in Microsoft® Excel®.

| | |
|---|---|
| | **Note:** Functions will be covered in more detail later. |

You can create a calculated field by right-clicking the **Data** pane and selecting **Create Calculated Field**, or by typing a formula directly in the **Rows** or **Columns** shelf. Calculated fields are shown with other fields in the **Data** pane and have an equal sign ( = ) in front of them.

You can create a calculated field named Profit by subtracting the product costs and shipping costs from sales based on data in the Products table. It would look like the following:

```
SUM[Product].[Sales] - SUM[Product].[Cost] - SUM [Product].[ShippingCost] =
[Profit]
```




SUM[Product].[Sales] – SUM[Product].[Cost] – SUM [Product].[ShippingCost] = [Profit]

*Figure 5-1: Calculated fields and calculations.*

| | |
|---|---|
| | **Note:** To learn more about using calculations, check out the Spotlight on **Best Practices for Using Calculations** presentation from the **Spotlight** tile on the CHOICE Course screen. |

#### Additional Information

For information about the types of calculations, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_understand_types.html**.

For information about creating calculated fields, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_formulas.html**.

For more information on when to use calculations, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_understand.html**.

## Calculation Editor

When you create a calculation from the **Data** pane by selecting **Create Calculated Field**, and when you edit a calculated field, the calculation editor will appear. The left side of the calculation editor has a field for you to name the calculated field. It's a good idea to name each calculated field descriptively so users will know what the calculation does. Below the name is the formula window where you can construct the formula by typing operators or field names, and by dragging and dropping fields from the **Data** pane.

The right side of the calculation editor has a function selector drop-down box where you can select functions you wish to use in your formula. To the right of the function selector, function help information about the purpose and syntax for the selected function appears. The help window on the right can be collapsed to save space in the calculation editor.

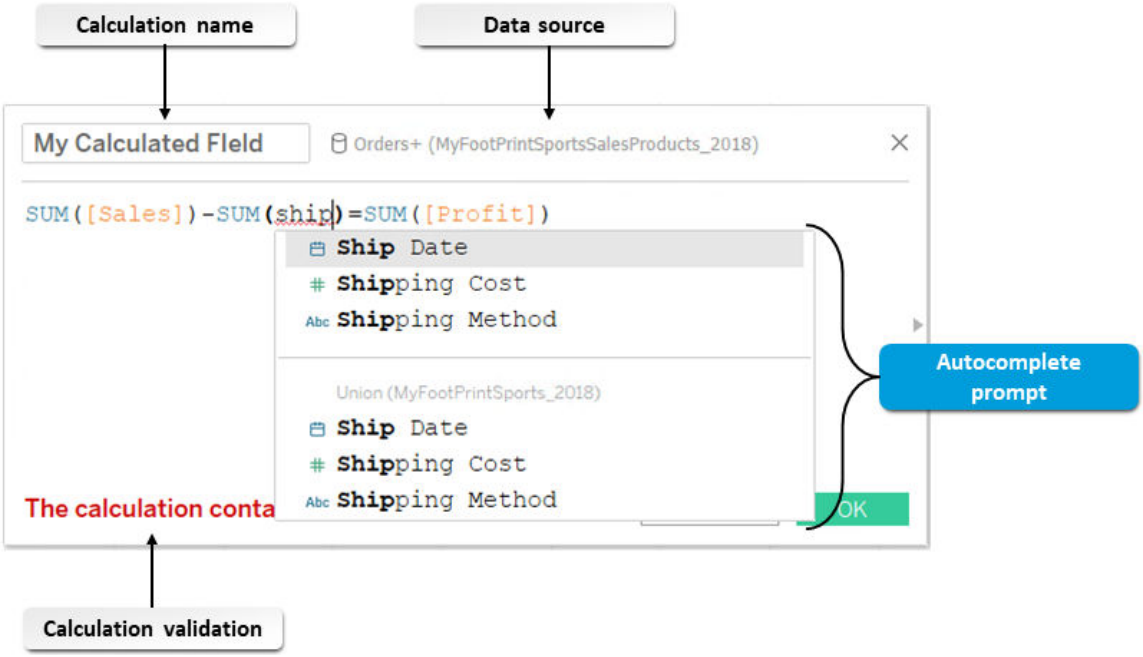



*Figure 5-2: The calculation editor.*

The formula window offers auto-complete to help you create formulas correctly. If Tableau identifies the function you're using, it will also display the appropriate function help information on the right side of the editor. You can drag-and-drop fields from the **Data** pane into your formulas as you're building them. The editor also color codes different portions of formulas. For example, operators are black, aggregations are blue, fields are orange, and parameters are purple. At the bottom-right of the formula window, the editor will display any errors it finds in your formula.

| | |
|---|---|
| | **Note:** To learn more about using the calculation editor, check out the Spotlight on **Working with the Calculation Editor** presentation from the **Spotlight** tile on the CHOICE Course screen. |

## In-Shelf Calculations

Tableau allows you to add calculations directly to a shelf. Simply start typing the calculation in the shelf. You can use auto-complete with in-shelf calculations, but you have to press **Enter** to select the correct auto-populated field. In-shelf calculations also support drag-and-drop additions of fields to build formulas.

Once completed, the calculation will appear in the pill in the shelf, and will be added to the view. Unlike calculated fields created from the **Data** pane, in-shelf calculations aren't automatically saved as a calculated field. To save an in-shelf calculation, drag the pill containing the in-shelf calculation to the **Data** pane and rename it, if you choose.

Valid in-shelf calculations are shown with a green pill. In-shelf calculations that contain errors are displayed with a red pill.

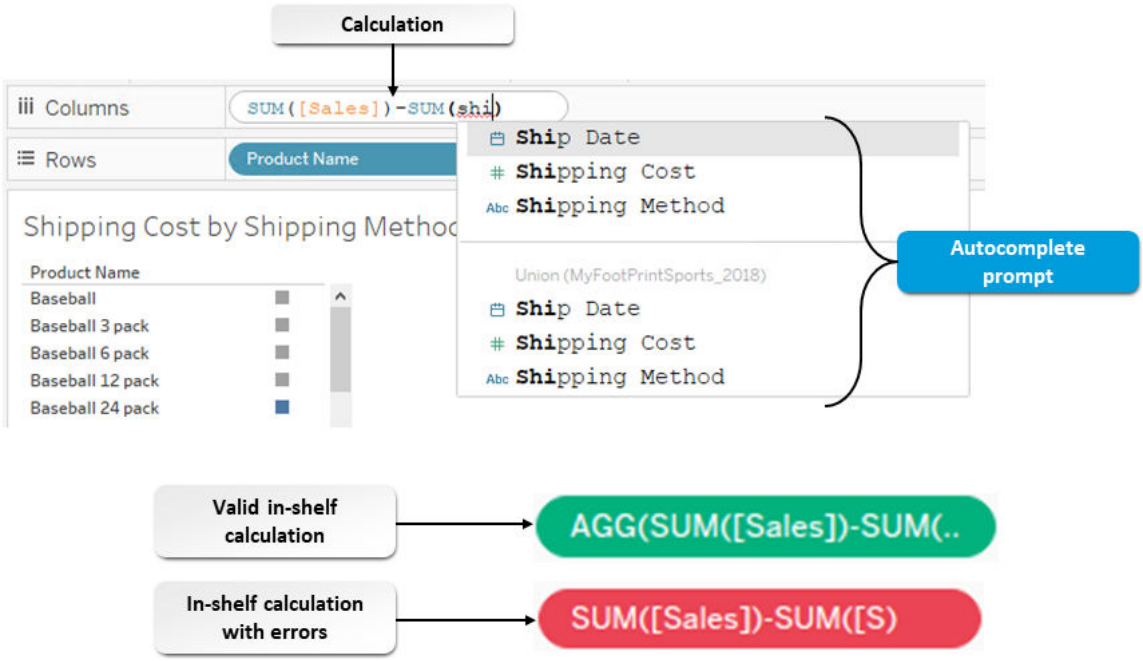



*Figure 5-3: An in-shelf calculation.*

### Additional Information

For more information on working with in-shelf calculations, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_adhoc.html**.

## Calculation Syntax for Calculated Fields

Calculations use the following syntax.

| *Component* | *Description* | *Examples* |
|---|---|---|
| **Comments** | Provides information about what the calculation or calculated field is for and how to use it. The line starts with two slashes //. | // This field calculates profit. |
| **Function** | Calls a Tableau function. The format is the function name followed by the expression in parentheses. Functions are colored blue in the calculation editor. | SUM (expression) |

| Component | Description | Examples |
|---|---|---|
| **Fields** | To include a field whose data you will use in the calculation, type the field name surrounded by brackets [ ] or drag-and-drop the field into the formula. Fields are colored orange in the calculation editor. | [Sales] |
| **Operators** | +, -, *, /, %, ==, =, >, <, >=, <=, !=, <>, ^, AND, OR, NOT, ( ) are used to build equations in calculated fields. | [Sales] – [Costs] |
| **Literal expressions** | Numeric; written as numbers. | 2.531 or 38 |
| | Strings must be surrounded by quotation marks. | "Profitable" or "Not Profitable" |
| | Dates are surrounded by hashtags ( # ). | #August 28, 2018# |
| | Booleans (true/false) are written as true or false. | true OR false |
| | Nulls are written as null. | Null |
| **Parameters** | To include a parameter in the calculation, type the parameter name surrounded by brackets [ ] or drag-and-drop the field into the formula. Fields are colored purple in the calculation editor. | [parameter] – [Costs] |

### Additional Information

For information about getting started with calculations, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_create.html**.

## Logic Calculations

Logic calculations are designed to determine if a condition expressed in a formula is true or false. This also called Boolean logic. IF-THEN-ELSE statements are an example of a logic calculation. Once you determine if a statement is true or false, you can take action based on the determination made by the logic formula.

To determine if specific stores are profitable, you could make a simple calculated field to subtract store expenses from store profit and return a string value. This calculation would use an IF-THEN-ELSE logic function that might look something like this:

```
IF SUM[Profit] > 0
THEN "Profitable"
ELSE "Unprofitable"
END
```

The first line takes the SUM of the Profit field and compares it using the Greater Than arithmetic operator. If the profit is greater than zero, the string value of "Profitable" is returned. String values must be encapsulated in quotations. If the profit is not greater than zero, the string value of "Unprofitable" is returned. The END statement is required to end the logical operation in the formula.

You could use this calculated field in views, such as ***dashboards*** to show when each store in a retail chain goes from unprofitable to profitable.

### Additional Information

For more information on the types of calculations you can create in Tableau, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_understand_types.html**.

## Guidelines for Creating Calculated Fields

| | |
|---|---|
| | **Note:** All of the Guidelines for this lesson are available as checklists from the **Checklist** tile on the CHOICE Course screen. |

### Creating Calculated Fields

When creating calculated fields, consider the following guidelines:

- While not required, it's recommended to put logic statements in all CAPS to make them easier to pick out in formulas.
- You don't need brackets for fields with a single word name, but you do need them if a field name contains any spaces.
- Watch the error indicator at the lower left of the formula window to make sure your formulas are correct.
- Remember that you can drag-and-drop fields from the **Data** pane into formulas in the calculation editor and to in-shelf calculations.
- Drag-and-drop all or part of a formula from the formula window to the **Data** pane to create a new calculated field based on the formula.
- Use the function helper for guidance when constructing formulas using functions.
- Use the **CRTL+** and **CTRL-** key combinations to increase or decrease font size in the formula window.

# ACTIVITY 5-1

## Analyzing Data with In-Shelf and Logical Expressions

### Data File

C:\095205Data\Manipulating Data with Calculations\MyFootPrintSports_2018_5.twb

### Before You Begin

Tableau Desktop is open.

### Scenario

You created a view that shows the profit by subcategory. You want to be able to see what the profit percentage is as well, but you don't want to further clutter the view with more data. A calculated field will allow you to calculate the profit percentage, and adding it to the tooltip won't add clutter to the view. You also want to see a map of sales by city that calls out cities that have $2,000 or more in sales. You will use a logic calculation to accomplish this.

---

1. Open the worksheet for the lesson.
   a) From the menu, select **File→Open**.
   b) In the **Open** dialog box, browse to the **C:\095205Data\Manipulating Data with Calculations** folder.
   c) Select **MyFootPrintSports_2018_5** and select **Open**.

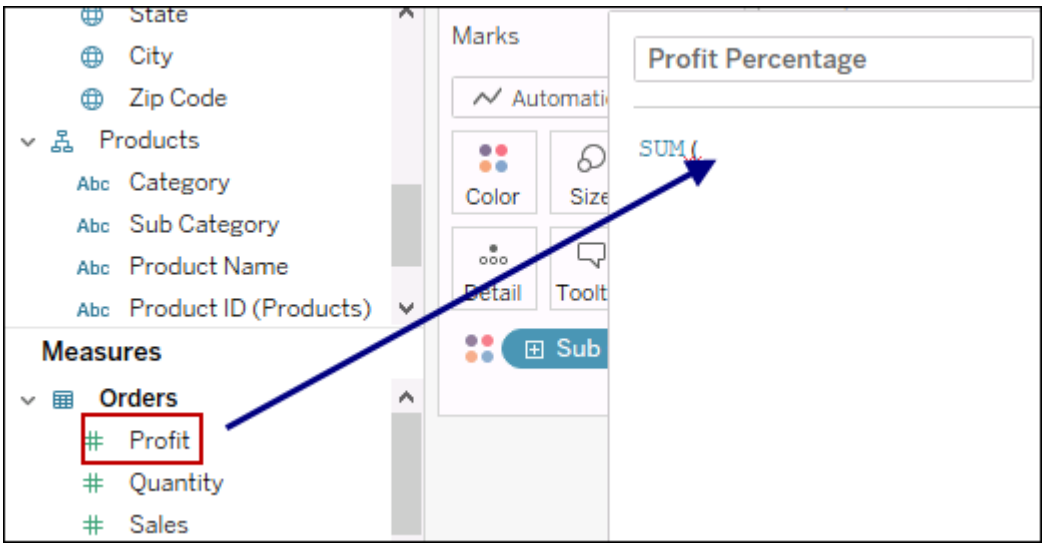
2. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Manipulating Data with Calculations** folder.
   c) In the **File name** box, type *MyCopy_MyFootPrintSports_2018_5* and select **Save**.

3. Create a calculated field that calculates the profit percentage.
   a) At the bottom of the screen, select the **Profit by Subcategory** worksheet.
   b) From the menu, select **Analysis→Create Calculated Field**.
   c) In the dialog box, in the **Calculation1** box, type *Profit Percentage*

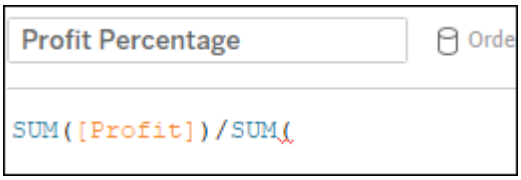
d) In the code area, type `SUM(`

| | |
|---|---|
| | **Note:** In this course, you can refer to the images and use those as a guide when typing calculation code. |
| | **Note:** As you type, the autofill will display options for what you are typing. |

e) In the **Data** pane, under **Measures→Orders**, select and drag **Profit** to the code area, at the end of the line.




f) In the code area, at the end of the line of code, type `)/SUM(`

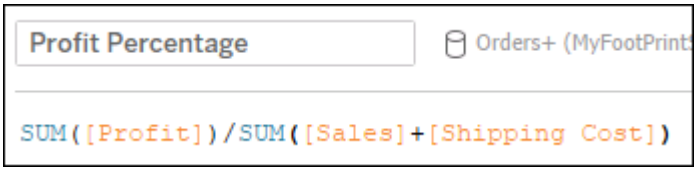



g) In the **Data** pane, under **Measures→Orders**, select and drag **Sales** to the code area, at the end of the line.

h) In the code area, at the end of the line of code, type `+`

i) In the **Data** pane, under **Measures→Orders**, select and drag **Shipping Cost** to the code area, at the end of the line.

j) In the code area, at the end of the line of code, type )




| | |
|---|---|
| | **Note:** At the bottom of the code area, Tableau will indicate if the calculation code is valid or invalid. If your code is invalid, the first places to check are ensuring brackets and parentheses are in the correct place. |

k) Select **OK**.

## 4. Format the calculated field.

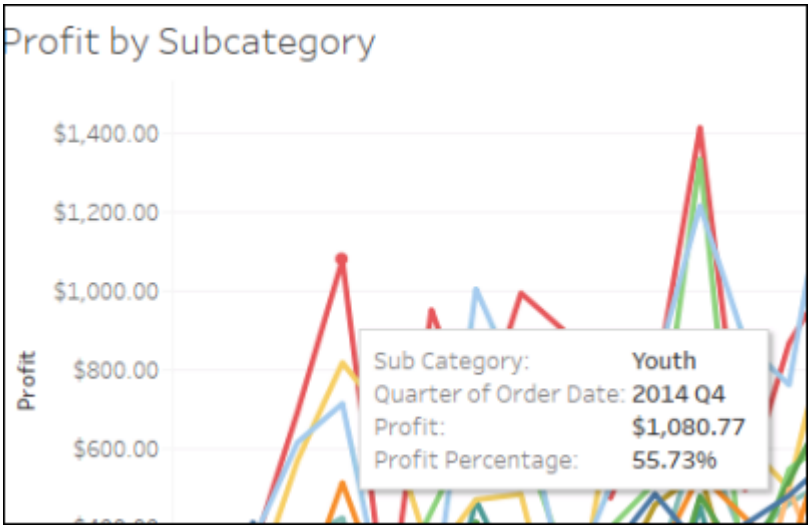
a) In the **Data** pane, under **Measures**, right-click **Profit Percentage**, and then select **Default Properties→Number Format**.

b) In the **Default Number Format** dialog box, select **Percentage** and select **OK**.

## 5. Add the calculated field to a view.

a) In the **Data** pane, under **Measures**, select and drag **Profit Percentage** to the **Tooltip** box in the **Marks** card.

b) In the canvas, hover the mouse over one of the lines and observe the profit percentage in the tooltip.




## 6. Create a view for sales by subcategory.

a) On the toolbar, select the **New Worksheet** button.

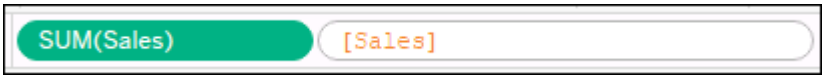
b) On the **Sheet 7** worksheet tab, in the **Data** pane, under **Dimensions→Products**, select and drag **Sub Category** to the **Rows** shelf.

c) In the **Data** pane, under **Measures→Orders**, select and drag **Sales** to the **Columns** shelf.

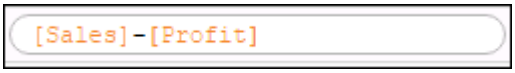
## 7. Add an in-shelf calculation that calculates cost.

a) In the **Columns** shelf, to the right of **SUM(Sales)**, double-click to open a blank pill.

b) Type `Sales` and then press **Enter**. Tableau will select the **Sales** field.




c) Type `-Profit` and then press **Enter**. Tableau will select the **Profit** field.




d) Press **Enter** again to create the calculated field.
e) In the canvas, observe the column for cost, which is sales minus profit.

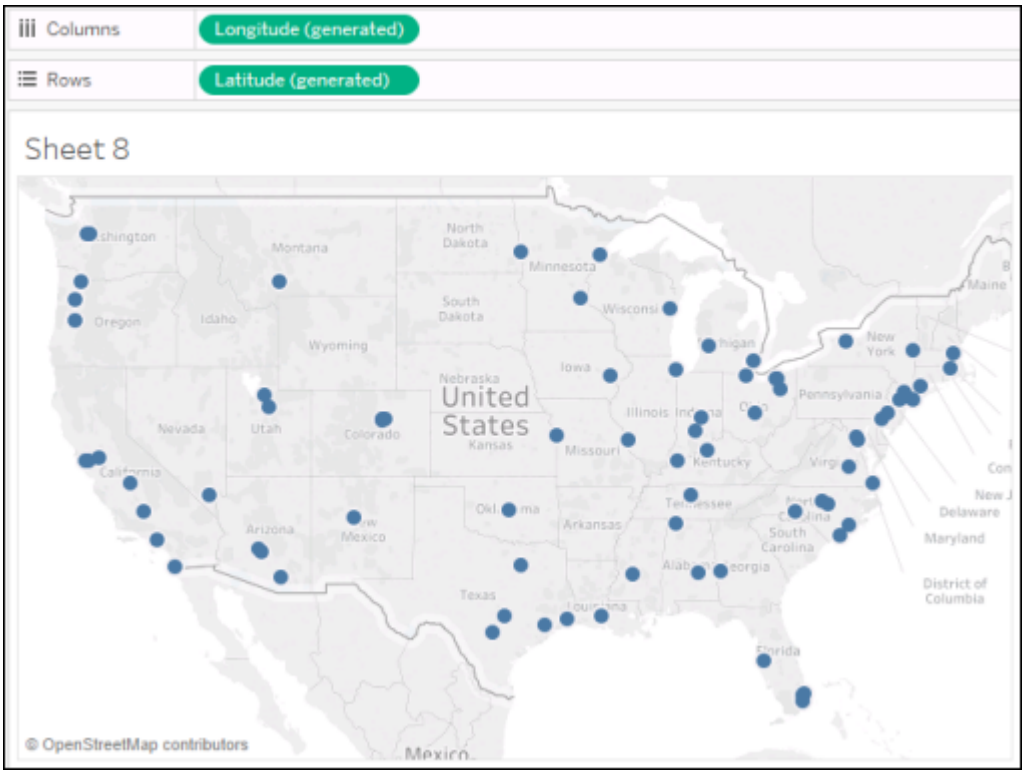
## 8. Rename the worksheet.

a) Right-click the **Sheet 7** tab and select **Rename**.
b) Type ***Sales and Cost by Subcategory*** and press **Enter**.

## 9. Create a map view.

a) On the toolbar, select the **New Worksheet** button.
b) On the **Sheet 8** worksheet tab, in the **Data** pane, under **Dimensions→Location**, double-click **City**.

c) In the canvas, observe how a map with the city locations was created.




d) In the **Data** pane, under **Measures→Orders**, select and drag **Sales** to the **Tooltip** box in the **Marks** card.

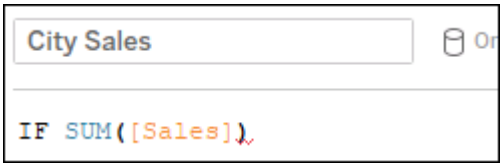
e) In the canvas, hover your mouse over some of the city marks to view the tooltip and the sales value.

10. Create a logic calculation that will color marks with $2,000 or more in sales in one color, and marks under $2,000 in another color.

a) From the menu, select **Analysis→Create Calculated Field**.

b) In the dialog box, in the text box, type ***City Sales***

c) In the code area, type `IF SUM([Sales])`




d) In the code area, at the end of the line of code, type `>=2000` and press **Enter**.

```
IF SUM([Sales])>=2000
```

e) In the code area, type `THEN "2000 or Above"` and press **Enter**.

```
IF SUM([Sales])>=2000
THEN "2000 or Above"
```

f) In the code area, type `ELSE "Below 2000"` and press **Enter**.

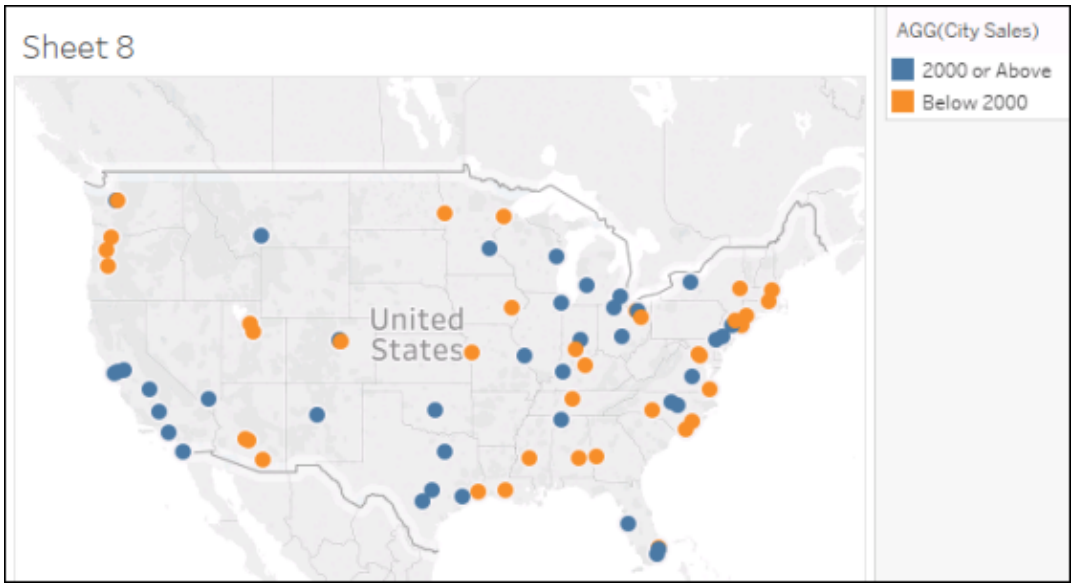
```
IF SUM([Sales])>=2000
THEN "2000 or Above"
ELSE "Below 2000"
```

g) In the code area, type `END`

```
IF SUM([Sales])>=2000
THEN "2000 or Above"
ELSE "Below 2000"
END
```

h) Select **OK**.

11. a) In the **Data** pane, under **Measures**, select and drag **City Sales** to the **Color** box in the **Marks** card.
    b) In the canvas, observe the cities are now in two groups. Those with $2,000 or more in sales and those with less than $2,000.




12. Rename the worksheet.
    a) Right-click the **Sheet 8** tab and select **Rename**.
    b) Type ***Sales by City Map*** and press **Enter**.

# TOPIC B

## Manipulate Data with Functions

Beyond calculating data, you may need to perform operations on data that fundamentally change or combine it. In this topic, you will manipulate data with calculations.

### Functions

Tableau supports a large number of functions that can be used in formulas. The following table lists the categories of functions and provides some of the examples of functions available in each category.

| *Function Category* | *Some Example Functions* |
|---|---|
| **Number functions** | Includes functions for working with numbers such as ZN to convert nulls to zeros and ROUND to round numbers to a specific decimal place, and arithmetic functions such as LOG for logarithm, SIN, and COS. |
| **String functions** | Includes functions such as CONTAINS to find a substring inside of a string, and REPLACE to replace one substring with another. |
| **Date functions** | Includes functions like DATEDIFF to find the length of time between two dates, and ISDATE to find a date that is part of a string. |
| **Logical functions** | Includes functions to help you determine if a given statement is true or false, such as IF-THEN-ELSE functions, and IFNULL to locate nulls. |
| **Aggregate functions** | Includes functions for aggregating values such as SUM, AVG to find the average, MIN and MAX to find minimum and maximum, and COUNT to find the number of rows in a data set. |
| **User functions** | Includes functions to return information about users logged into Tableau Server and Tableau Online such as USERNAME to find the user ID of the current user, and USERDOMAIN to find the domain of the current USER. |
| **Table calculations** | These functions to be used to retrieve information about data stored in tables. |
| **Pass-through functions (RAWSQL)** | These functions can send Structured Query Language (SQL) expressions directly to the data source database without Tableau trying to interpret them first. These are typically used to call custom database functions that Tableau is not aware of. |
| **Additional functions** | Includes regular expressions, Hadoop® Hive-specific functions, and Google BigQuery™-specific functions that you can call from Tableau. |

#### All Tableau Functions by Category

For information on functions available in Tableau by category, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/functions_all_categories.html**.

### Number Calculation Examples

Numeric calculations compute numeric fields from your data source and show the results. There are many ways raw data can be enhanced by creating numeric calculated fields. For example, if you wanted to see the profit for each product, but only have data for gross sales, and product cost, and

shipping expenses, you might create a calculated field named [Profit] based on the following formula:

```
SUM([Sales]-[Cost]-[Shipping Cost])= SUM([Profit])
```

You can also perform data cleanup using calculations, for example, if a field has many null values after joining or unioning different tables. Or, if you have nulls for products that are sold in-store only and don't have a shipping cost. Since a null value will yield a result of null in a calculation, you have to manage the nulls. The ZN function was designed to convert null values to zeros so computations can be performed. You can use ZN with the aforementioned calculation in the following way:

```
ZN(SUM([Sales]-[Cost]-[Shipping Cost])= SUM([Profit]))
```

## Date Calculation Examples

Date calculations let you determine dates through calculations, organize data by date, and do things like determine the length of time between events. Data functions can use the Gregorian or the ISO 8601 Standard calendars. For example, if you wanted to see how quickly your invoices are being paid by your corporate customers to predict cash flow, you could create a calculation like this:

```
DATEDIFF("day" [invoice date],[paid date])
```

In this case, the DATEDIFF function is used to find the difference between two dates. You must specify the part of the date you want the result to be enumerated in. You can select day, month, quarter, or year; it must be lowercase and surrounded by quotes. Then the DATEDIFF function calculates the number days between the dates from the [invoice date] field and the [paid date] field.

Another date-related calculation might be to separate orders by quarters to see which quarters are the best, or to calculate quarterly numbers or sales commissions. To do this, you might use a calculation like the following:

```
DATETRUNC ("quarter", [Sale Date])
```

DATETRUNC truncates a date by the date part specified and returns a new date. If you specified a date part of "month," then a [Sale Date] of March 15 would generate a result of March 1. In this case, the date part is "quarter," so all sales that take place between January 1 and March 31 of the year generate a result of 1/1/<Year>. Dates that fall between April 1 and June 30 generate a result of 4/1/<Year>. In this way, the calculated field can be used to segment orders by quarter.

## String Calculation Examples

String calculations are useful for finding sub-strings of text in string fields, replacing sub-strings, splitting strings, and extracting specific portions of string data. For example, if you wanted to find out how many fans were sold over the summer months, you could search sales record for any product described as a fan. The calculation might look something like this:

```
CONTAINS([Product Description],[fan])
```

The CONTAIN function will return true if the sub-string, in this case fan, is found in the [Product Description] field.

One of the most commonly used string functions splits strings into their component parts. This function is used so commonly it's available from the menu. You could use this function to split a name field into multiple parts. For example, if your data source had a name field that contained both the first and the last name separated by a space, you could split that string as a part of a calculated field to extract the first name:

```
SPLIT([Name], ' ', 1)
```

This calculation uses the SPLIT function to split a string in the [Name] field using a space as a delimiter. You must also specify a token. Each separate part of the string that is split is a token. In this case, each field will have two tokens: the text prior to the delimiter, which is the first name, and

the text after the delimiter, which is the second name. The token specified is 1 because the goal was to extract the first name.

You can also use string calculations to extract specific parts of a string field. For example, if the Order Number field is generated by the point of sale system and starts the eight-digit Product ID of the product sold, you can use a calculation to extract the Product ID from Order Number:

```
LEFT([Order Number], 8)
```

The LEFT function starts at the left of the string in the [Order Number] field and returns the first eight characters in the string from the left, which is the Product ID you're looking for.

## Example Calculations Where Type Change is Required

There may be instances where the type of data stored in a field should be changed to make the data easier to compute, analyze, or understand. For example, the first six characters of each Order ID may contain the order date. If Order ID is a string field, you can use the LEFT function to extract the order date. If, however, Order ID is a numeric field, you must convert the field to a string using a calculation like the following:

```
LEFT(STR([Order ID]),6)
```

The LEFT function will capture the first six characters of the [Order ID] field after the STR function converts the field to a string.

In another example, you may wish to reward any customers that have been customers for longer than 10 years. You could do this with the following calculation:

```
DATEDIFF('day', [Account Created Date],TODAY())/365
```

This calculation uses the DATEDIFF function to find the days since the account was created. It does this by finding the difference in dates between the time the account was created and the current date, and then dividing the result by 365 to get the number of years. The problem is, by default, this yields a decimal number such as 8.5, or 11.3 when you just want whole numbers. To get whole numbers, you can convert the entire equation to integers in the following way:

```
INT(DATEDIFF('day', [Account Created Date],TODAY())/365)
```

By wrapping the entire formula in the INT() function, the result is converted to an integer.

**Note:** This formula uses the day datepart instead of the year datepart because using the year would result in rounding errors that could throw off the results by a year.

## Aggregate Calculation Examples

There may be times when you need to change the granularity of the data you're working with. For example, if you're analyzing profit, you may have developed the following calculation to show the profit for each order:

```
SUM([Order ID]-[Cost]-[Shipping Cost])
```

If you wanted to see the average profit per order, you could build aggregation into the calculation in the following way:

```
AVG([Order ID]-[Cost]-[Shipping Cost])
```

Normally, aggregation is performed after a field is added to the view. However, if aggregation is built into the calculation, as in the aforementioned example, aggregation will occur as specified as part of the calculation. This guarantees the level of granularity defined by the aggregation. Pills that have AGG at the beginning of the name have aggregation done as part of the calculation.

### Aggregate Functions in Tableau

You already know that measures are routinely aggregated in Tableau to show relevant data at a defined level of granularity. Aggregations might show the SUM of sales data for a store, the

AVERAGE number of products sold annually, or the MEDIAN price of product sold. You can also use aggregation functions in calculations to change the granularity of the data or summarize it.

The aggregation functions available in Tableau are as follows.

| *Aggregation* | *Description* |
|---|---|
| ATTR | Attribute returns the value of an expression if there is only a single value from all computed rows. If there are multiple values, it returns an asterisk. It ignores any null values. |
| AVG | Average returns the average of all computed rows. This can be used with numeric fields only. It ignores any null values. |
| COLLECT | Collection is used with spatial fields to combine values. |
| CORR | Correlation returns the linear relationship between two expressions. |
| COUNT | Count returns the number items, such as rows, in the dimension or measure. Numbers, dates, booleans, and strings can be counted. It ignores any null values. For example, if a table had 300 rows total, COUNT would return the number 300. |
| COUNTD | Count distinct returns the number of unique values in the dimension or measure. For example, if a table had 300 rows total, with 117 rows that contained unique data, COUNTD would return 117. |
| COVAP | Covariance of population returns the population covariance of two expressions, showing how the pair has changed together. |
| MAX | Maximum returns the largest number of an expression based on the data computed. If computed values are strings, it returns the last alphabetical string value. |
| MEDIAN | Returns the median number of an expression based on the data computed. This can only be used on numeric fields. It ignores any null values. |
| MIN | Minimum returns the smallest number of an expression based on the data computed. If computed values are strings, it returns the first alphabetical string value. |
| PERCENTILE | Returns the percentile value from an expression corresponding to a specified number between 0 and 1 (inclusive). |
| STDEV | Standard deviation returns the standard deviation of all values in the given expression based on a sample population. Use this option if your data represents a sample of the population. |
| STDEVP | Standard deviation population returns the standard deviation of all values in the given expression based on a biased population. Use this for large sample sizes. |
| SUM | Returns the sum of all values in the expression. This can only be used with numeric fields. It ignores any null values. |
| VAR | Variance returns the variance of all values in the given expression based on a sample population. Use this option if your data represents a sample of the population. |

| *Aggregation* | *Description* |
|---|---|
| VARP | Shows the variance of all values in the given expression based on a biased population. Use this for large sample sizes. |

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Explore Data with Calculations.**

# ACTIVITY 5-2
## Exploring Data with Calculations

### Before You Begin

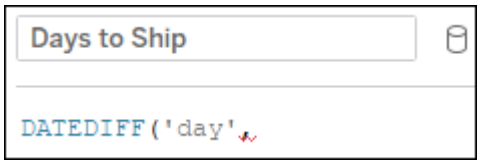
The **MyCopy_MyFootPrintSports_2018_5** workbook is open in Tableau Desktop.

### Scenario

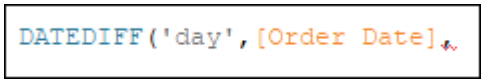
You want to know the average number of days it takes for an order to ship once it has been placed. You want to round off the number of days to whole days. There has been an incident involving some products shipped to an area of California. All of the locations have a ZIP Code that begins with 900. You want to isolate those orders in a view using a string calculation.

---

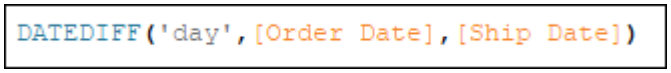
1. Create a date calculation to calculate the number of days it takes to ship an order.
   a) At the bottom of the screen, verify the **Sales by City Map** tab is selected.
   b) From the menu, select **Analysis→Create Calculated Field**.
   c) In the dialog box, in the text box, type ***Days to Ship***
   d) In the code area, type `DATEDIFF('day',`




   e) In the **Data** pane, under **Dimensions→Orders**, select and drag **Order Date** to the code area, at the end of the line.
   f) In the code area, at the end of the line of code, type `,`



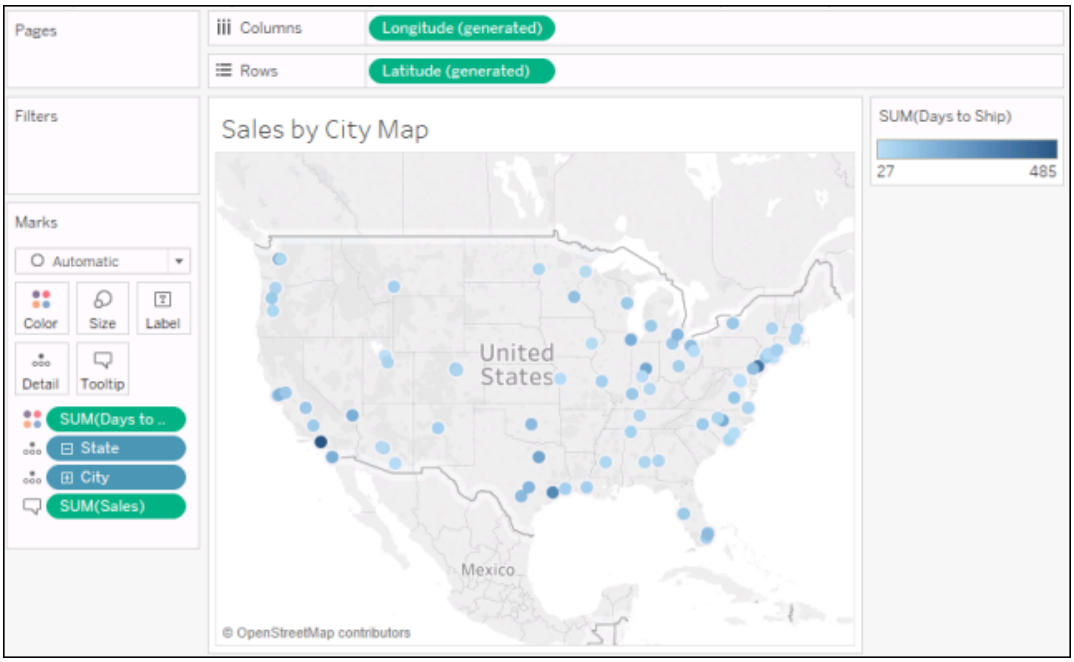

   g) In the **Data** pane, under **Dimensions→Orders**, select and drag **Ship Date** to the code area, at the end of the line.
   h) In the code area, at the end of the line of code, type `)`




   i) Select **OK**.

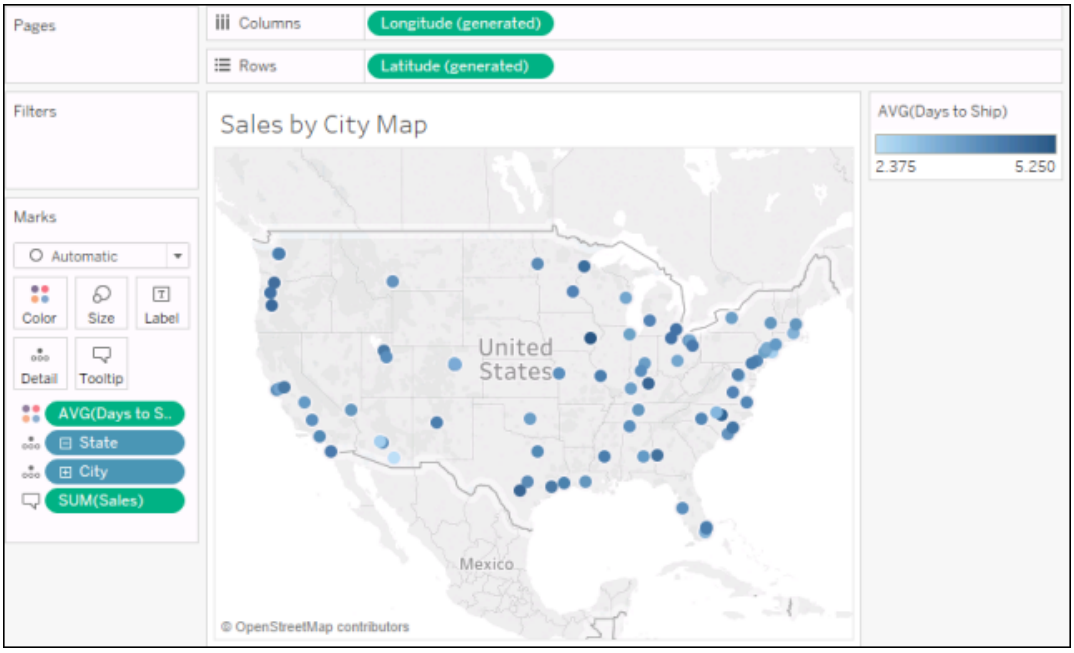
2. Use the date calculation.
   a) In the **Data** pane, under **Measures**, select and drag **Days to Ship** to the **Color** box in the **Marks** card.

b) In the canvas, in the **Legend**, in the **SUM(Days to Ship)** card, observe that the **Days to Ship** calculated field is the sum of all of the days to ship to a city, not the average.




c) In the **Marks** card, right-click **SUM(Days to Ship)** and select **Measure (Sum)→Average**.

d) In the canvas, observe that the **Days to Ship** calculated field is now showing the average days to ship.



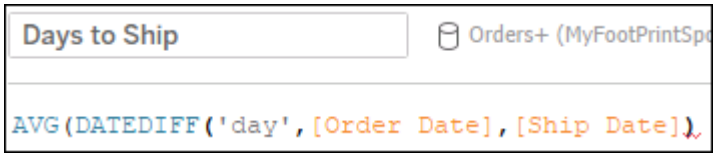

## 3. Add the average attribution to the date calculation.

a) In the **Marks** card, right-click **AVG(Days to Ship)** and select **Remove**.

You are removing this now because, when you add the AVG attribution to the code, this current pill will no longer function correctly and will cause problems in the map.

b) In the **Data** pane, under **Measures**, right-click **Days to Ship** and select **Edit**.

c) In the **Days to Ship** dialog box, in the code area, at the beginning of the line of code, type `AVG(`



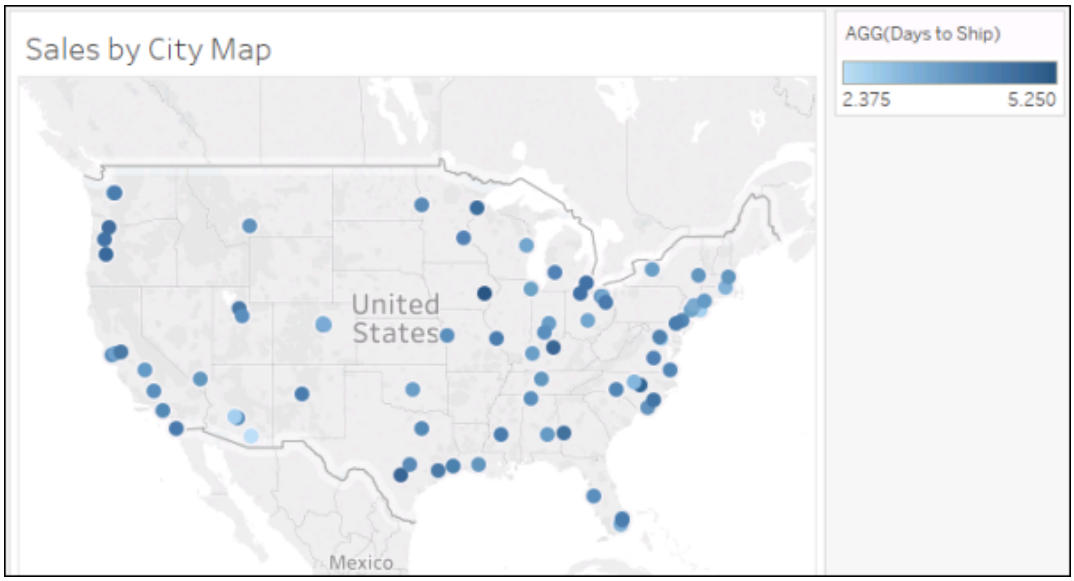

d) At the end of the line of code type `)`

```
AVG(DATEDIFF('day',[Order Date],[Ship Date]))
```

e) Select **OK**.

f) In the **Data** pane, under **Measures**, select and drag **Days to Ship** to the **Color** box in the **Marks** card.

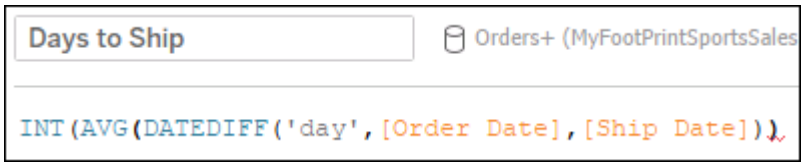
g) In the canvas, observe that the **Days to Ship** calculated field is showing the average days to ship.




4. Edit the calculated field to add the integer function to round off decimal places.

a) In the **Data** pane, under **Measures**, right-click **Days to Ship** and select **Edit**.

b) In the **Days to Ship** dialog box, in the code area, at the beginning of the line of code, type `INT(`

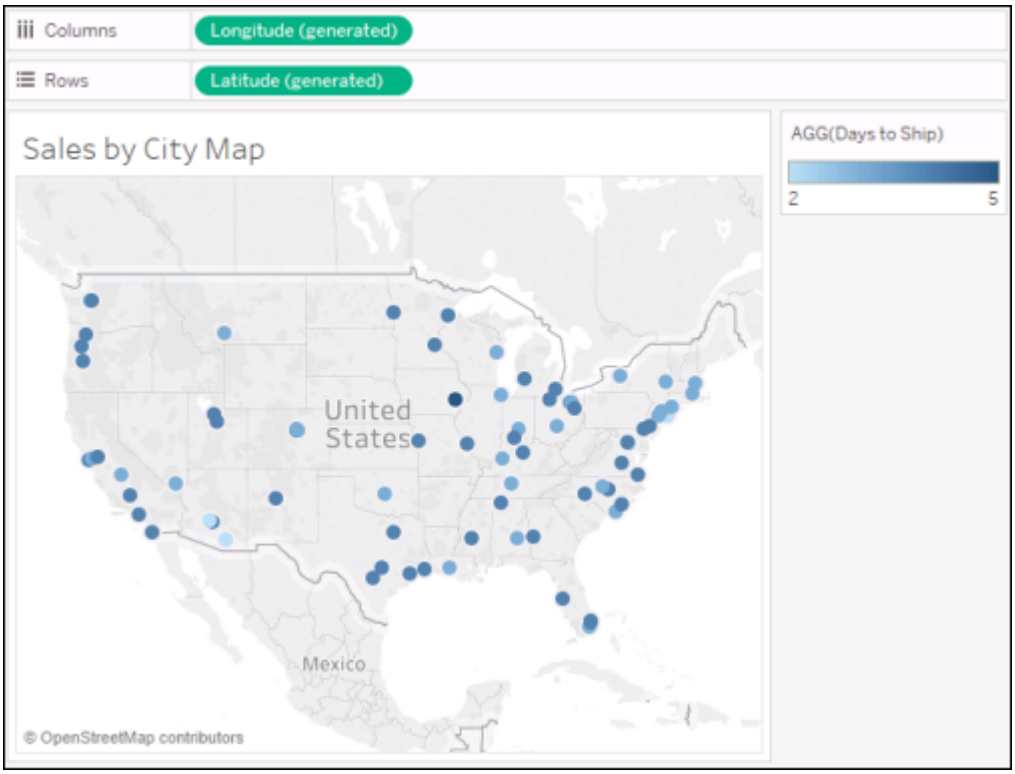



c) At the end of the line of code type `)`

```
INT(AVG(DATEDIFF('day',[Order Date],[Ship Date])))
```

d) Select **OK**.

e) In the canvas, observe that the **Days to Ship** calculated field is now rounded to the nearest whole number.

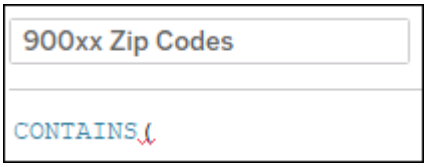



## 5. Create string calculation that selects ZIP Codes that contain 900.

a) From the menu, select **Analysis→Create Calculated Field**.

b) In the dialog box, in the text box, type ***900xx Zip Codes***

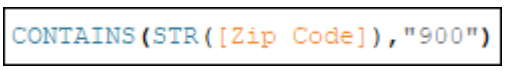
c) In the code area, type `CONTAINS(`




d) In the code area, at the end of the line of code, type `STR(` to convert the ZIP Code to a string value.

e) In the **Data** pane, under **Dimensions→Location**, select and drag **Zip Code** to the code area, at the end of the line.

f) In the code area, at the end of the line of code, type `),"900")`

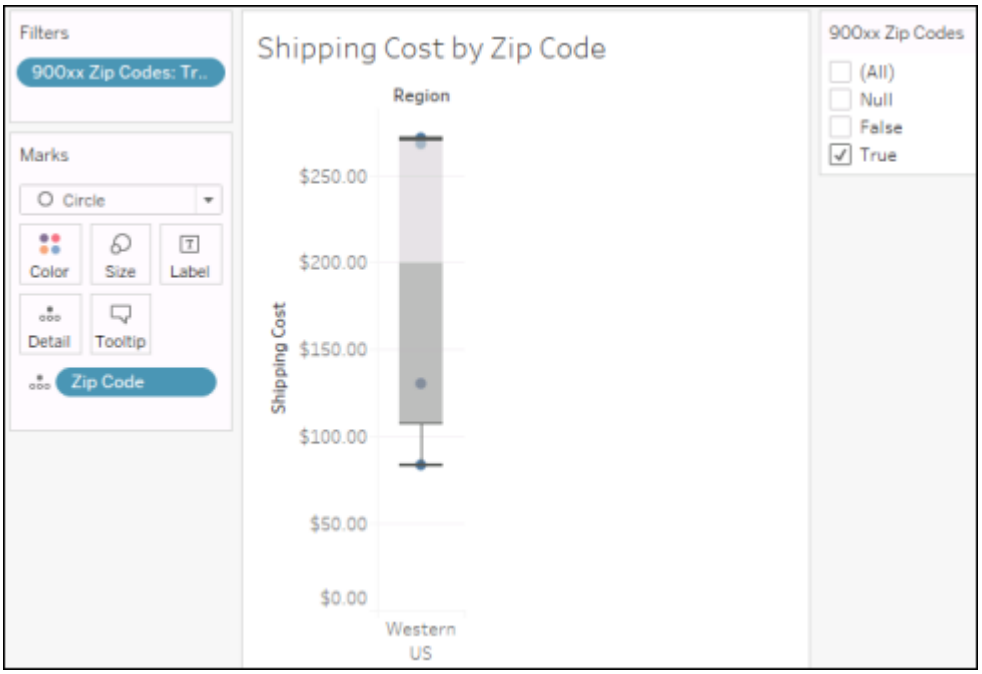



g) Select **OK**.

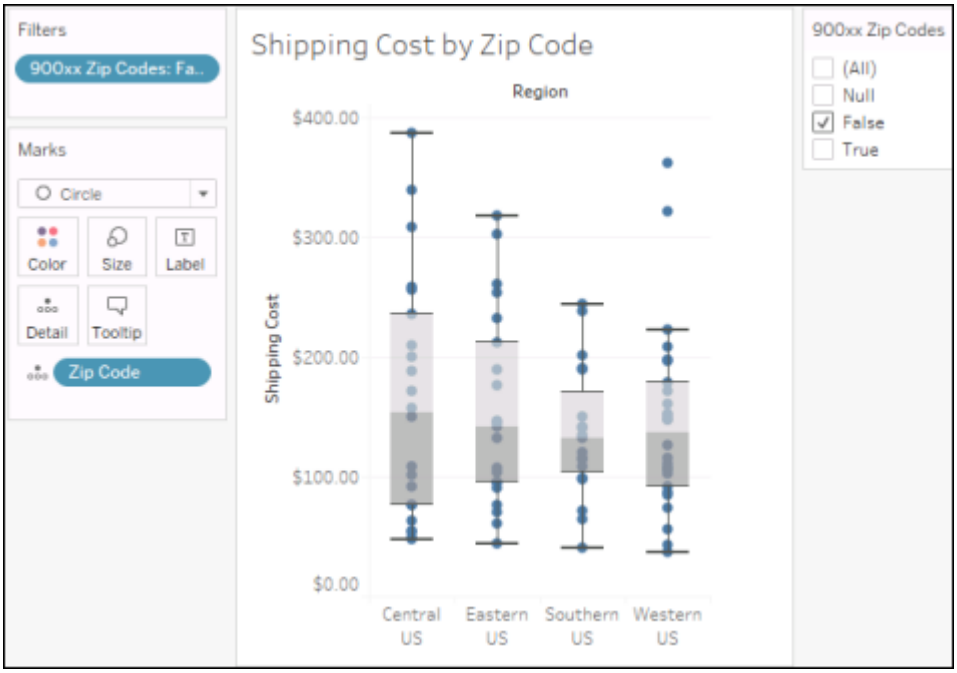
## 6. Use the string calculation to filter products displayed in a view.

a) At the bottom of the screen, select the **Shipping Cost by Zip Code** worksheet.

b) In the **Data** pane, under **Dimensions**, select and drag **900xx Zip Codes** to the **Filters** card.
c) In the **Filter [900xx Zip Codes]** dialog box, in the list, check the **True** check box.
d) Select **OK**.
e) In the **Filters** card, right-click **900xx Zip Codes: True** and select **Show Filter**.
f) In the legend, in the **900xx Zip Codes** card, verify the **True** check box is checked.
g) In the canvas, observe that only the three ZIP Codes that match the string calculation are displayed.




h) In the legend, in the **900xx Zip Codes** card, check the **False** check box and then uncheck the **True** check box.
i) In the canvas, observe that now all of the non-900xx ZIP Codes are displayed.



# TOPIC C

## Analyze Data with Table Calculations

In addition to performing calculations and manipulating data, some types of calculations can help you view data from a different perspective and help bring out different insights. In this topic, you will analyze data with table calculations.

### Table Calculations

In Tableau, regular calculations that you configure, such as [Sales] – [Expenses], are passed to the data source as a query. The data source computes the calculation and responds with the result. ***Table calculations*** are a special type of calculated field in Tableau. They are secondary calculations computed locally in Tableau on the data displayed in the visualization. These calculations don't use any dimensions or measures that have been filtered out of the visualization. Any field that has had a table calculation applied will have a triangle on the right side of the pill. They help you visualize data in different ways to meet your analysis needs.

Quick table calculations can be applied directly from the right-click menu, but if you need a custom table calculation, you can create one using the calculation editor by selecting **Table Calculation** from the list of functions in the editor. When creating a table calculation in the calculation editor, you must specify the type of aggregation to be used in the calculation.

For example, you can view percentage relationships, year-to-date totals and differences. As you change the calculation, the visualization updates. For example, for a sales measure that shows dollar amounts of products sold, you can perform a table calculation that shows percentages. That will update sales data displayed from dollar values in the visualization to show a percentage of total.

### Quick Table Calculations

***Quick table calculations*** are frequently used table calculations that you can apply to your visualization in Tableau. They are available from the right-click menu for fields, and are applied with the most common configuration for the calculation. Quick table calculations provide a quick way to visualize data differently, without having to take the time to manually build a table calculation.

You can perform the following types of quick table calculations in Tableau:

- **Running total**: A running total updates as new entries are added to the fields being totaled.
- **Difference**: Shows the difference between values for comparison purposes.
- **Percent difference**: Shows the difference between values as a percentage.
- **Percent of total**: Shows individual entries as a percentage of the total.
- **Rank**: Ranks values in the visualization.
- **Percentile**: Shows values as a percentage.
- **Moving average**: Shows data as an average over a range of time, filtering out uncommon fluctuations.
- **YTD total**: Shows totals from the start of the year to the current date.
- **Compound growth rate**: Shows growth over multiple time periods.
- **Year over year growth**: Shows growth relative to the last yearly time period.
- **YTD growth**: The change, positive or negative, from the similar period the previous year.

Quick table calculations are easy to create with a single click. You can then drag pills with a quick table calculation applied to the **Data** pane to save it as a new calculated field.

## Field Set

Table calculations are performed on the data in the ***field set***. The field set is the information in the view, and nothing that has been filtered out of the view. Essentially, the field set is the data behind the marks displayed in the view. One way to think of a field set is as if it were a simple Microsoft® Excel® spreadsheet. If you were to make a spreadsheet of your data, the rows and columns in the spreadsheet are what Tableau is using to perform the table calculation. This may be hard to conceptualize because many views look radically different than spreadsheets.

Because the field set is defined by the data that makes up the view, if you add or remove fields, or apply filters that change the view, you also change the field set. This alters the table calculation.

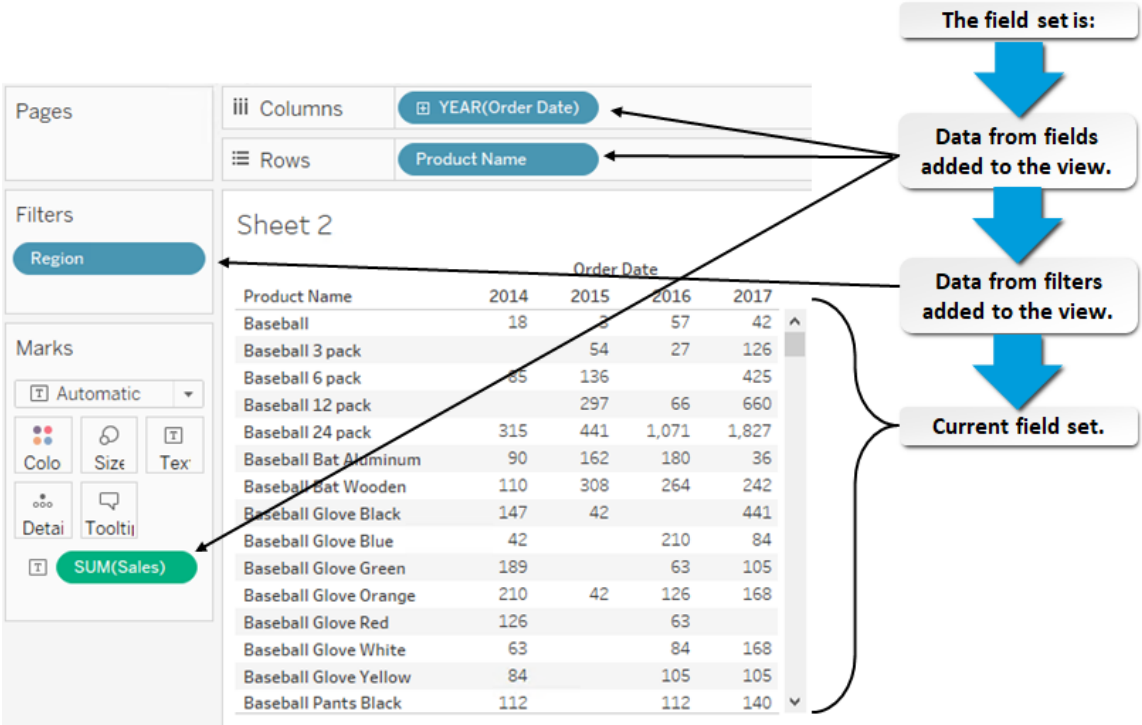



*Figure 5-4: A field set.*

## Partitions and Addresses

When computing a table calculation, Tableau uses the field set. But how does it perform the computation? There is more than one way to approach that. Think of it as if you're looking at a spreadsheet you're going to manually compute. Should you compute the entire table? Should you compute each row or column of data individually? Should you go across the rows then down? How you approach the computation affects the results of the computation.

How the data in the field set is ***partitioned*** and ***addressed*** defines how the computation is performed. The partition can be thought of as the scope of the calculation, or what will be computed. Partitions place data into groups, such as each pane, each row or column, or the entire table or spreadsheet. The partition tells Tableau what to include in the computation, when to start computing, and when to stop.

For example, if you're looking at sales data, you have the option to compute the entire table and see numbers for the entire year, or to compute each pane that represents a quarter and see quarterly results.

Addressing is similar to the direction of the computation. Again, think of a spreadsheet. The addressing tells Tableau how to perform the computation. For example, to compute moving across the table, or down the table.

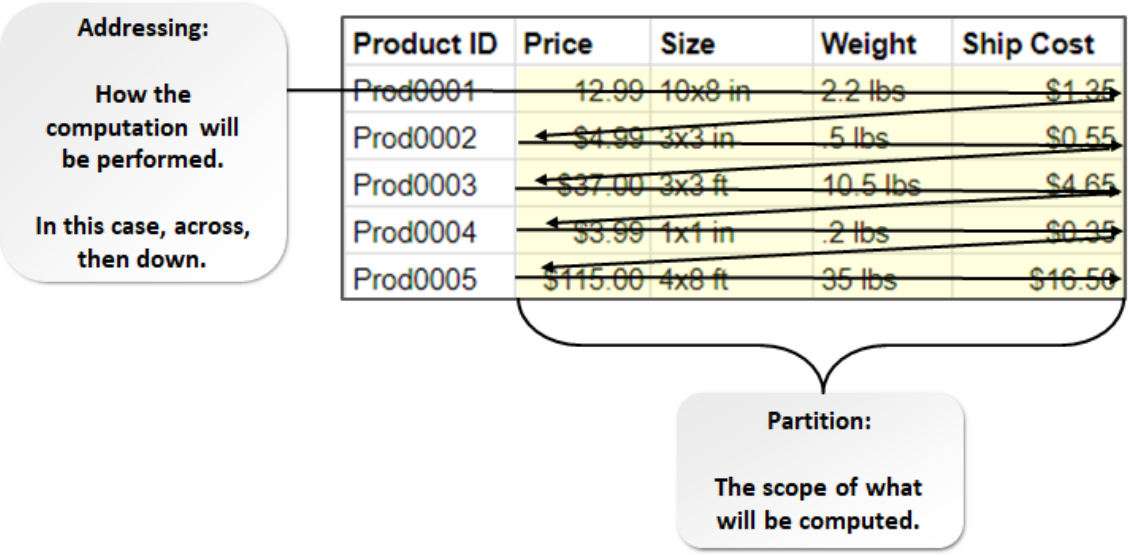



*Figure 5-5: Partitions and addresses.*

## The Compute Using Feature

The partitions and addresses used in table calculations are chosen for you by default, but you can adjust them by right-clicking the pill and selecting **Compute Using**. This allows you to select the method of partitioning and addressing you wish, and even select individual dimensions. This is useful if you know exactly how you want your table calculation performed; however, you may wish to experiment with different partitioning and addressing algorithms and see the results. To do this, select **Edit Table Calculation**. When you do, the dialog box that appears gives you the same **Compute Using** options and will update to reflect the new computation when it's selected. This gives you the opportunity to confirm you are computing the table calculation in a way that will give you the desired results.

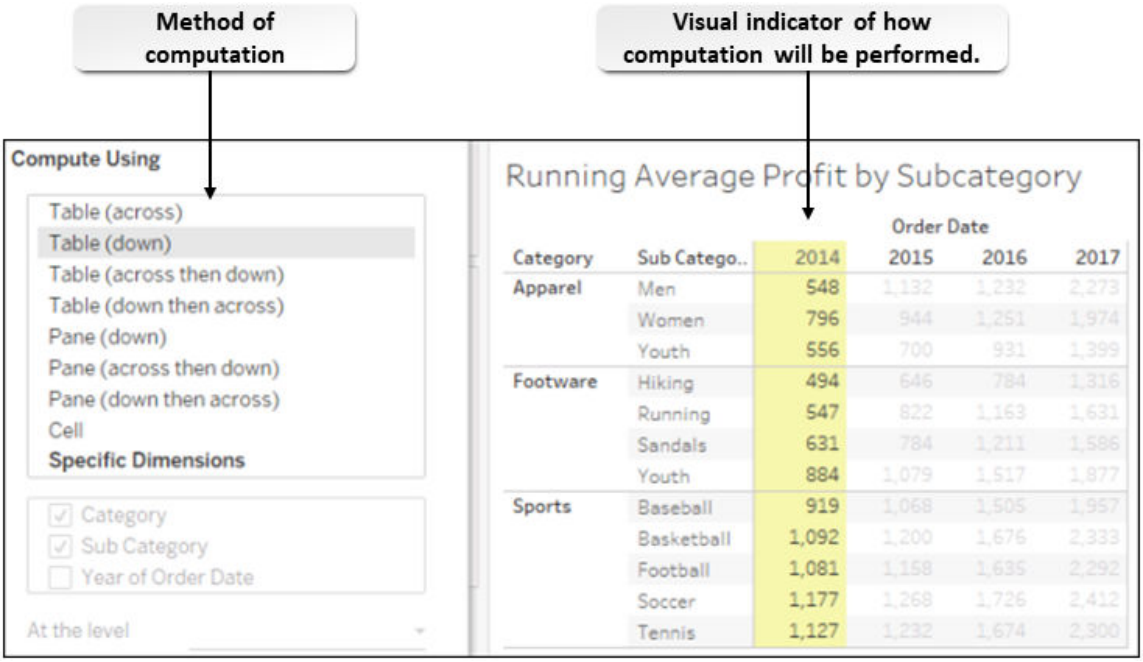



*Figure 5-6: The Compute Using feature.*

You can compute table calculations using the following options.

| *Compute Using* | *Description* |
|---|---|
| **Table (across)** | Processes across the length of the entire table and restarts at each partition. |
| **Table (down)** | Processes down the entire table and restarts at each partition. |
| **Table (across then down)** | Processes across the length of the entire table, then down the length of the table. This is similar to a table-wide, row-by-row approach. |
| **Table (down then across)** | Processes across the length of the entire table, then down the length of the table. This is similar to a table-wide, column-by-column approach. |
| **Pane (down)** | Processes down an entire pane. |
| **Pane (across then down)** | Processes across the pane, then down the pane. This is similar to a row-by-row approach within the pane. |
| **Pane (down then across)** | Processes down the pane, then across the pane. This is similar to a column-by-column approach within the pane. |
| **Cell** | Processes within a single cell. |
| **Specific Dimensions** | Processes using only the specific dimensions you check. |

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Analyze Data with Table Calculations.**

# ACTIVITY 5-3

## Analyzing Data with Table Calculations

### Before You Begin

The **MyCopy_MyFootPrintSports_2018_5** workbook is open in Tableau Desktop.

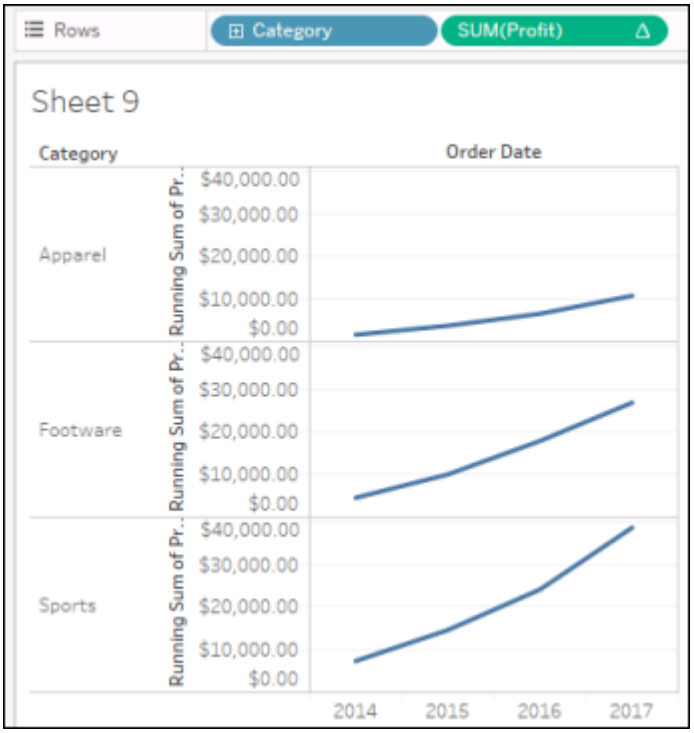
### Scenario

You want to see the running average for profit over the last four years for each category, and possibly each subcategory. You will need to create a table calculation for the running average and then explore different compute options. You want to compute across the years and then down the subcategories when performing the running average calculation.

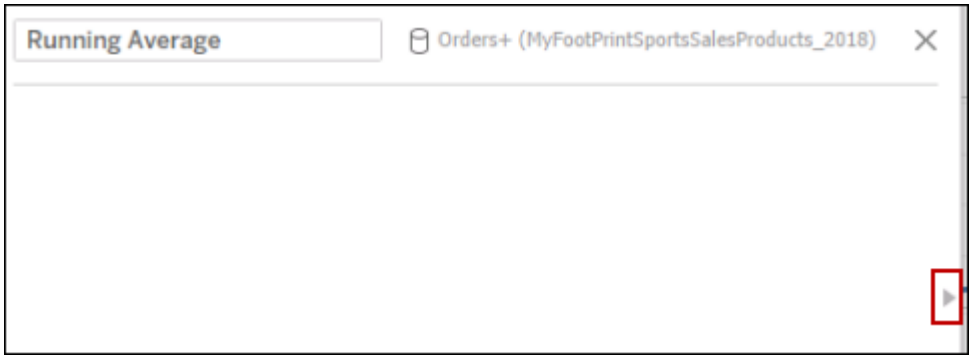
---

1. Create a view.
   a) On the toolbar, select the **New Worksheet** button.
   b) On the **Sheet 9** worksheet tab, in the **Data** pane, under **Dimensions→Orders**, select and drag **Order Date** to the **Columns** shelf.
   c) In the **Data** pane, under **Dimensions→Products**, select and drag **Category** to the **Rows** shelf.
   d) In the **Data** pane, under **Measures→Orders**, select and drag **Profit** to the **Rows** shelf.

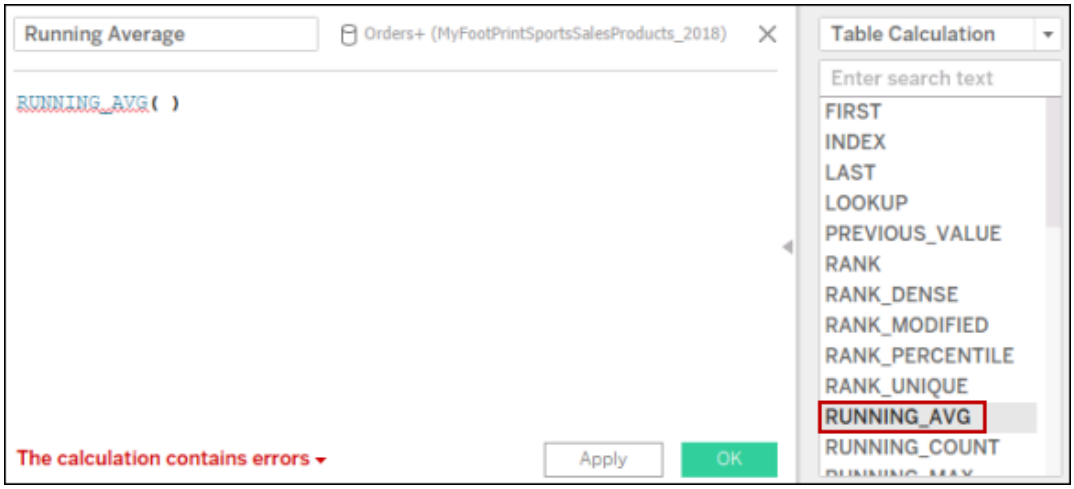
2. Add a quick table calculation for the running total.
   a) In the **Rows** shelf, right-click the **SUM(Profit)** pill and select **Quick Table Calculation→Running Total**.
   b) In the canvas, observe the new values for years **2015-2017** and notice the **Sports** category has the most total profit.



3. Create a table calculation that computes the running average for profit.
   a) From the menu, select **Analysis→Create Calculated Field**.
   b) In the dialog box, in the text box, type ***Running Average***
   c) In the code area, select the right arrow to expand the functions section.




   d) In the **All** drop-down list, select **Table Calculation**.
   e) In the list, double-click **RUNNING_AVG** to add it to the code area. This function will compute the running average for the measure you add to it.




   f) In the code area, at the end of the line of code, delete the closing parenthesis.
   g) In the code area, at the end of the line of code, type `SUM(`

```
RUNNING_AVG(SUM(
```

   h) In the **Data** pane, under **Measures→Orders**, select and drag **Profit** to the code area, at the end of the line.
   i) If there is a space before or after **[Profit]**, delete the space.
   j) In the code area, and the end of the line of code, type `))`

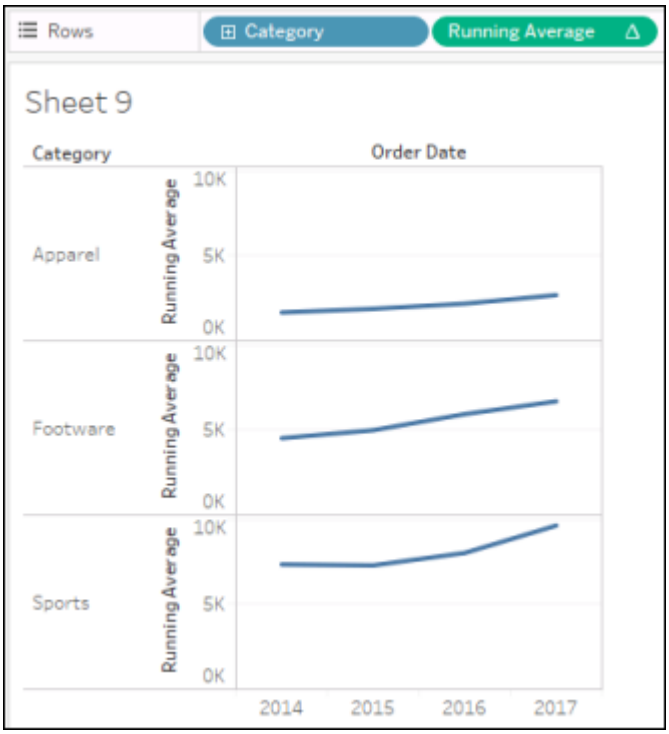
```
RUNNING_AVG(SUM([Profit]))
```

   k) Select **OK**.

4. Add the running average table calculation to a view.
   a) In the canvas, in the **Rows** shelf, right-click the **SUM(Profit)** pill and select **Remove**.

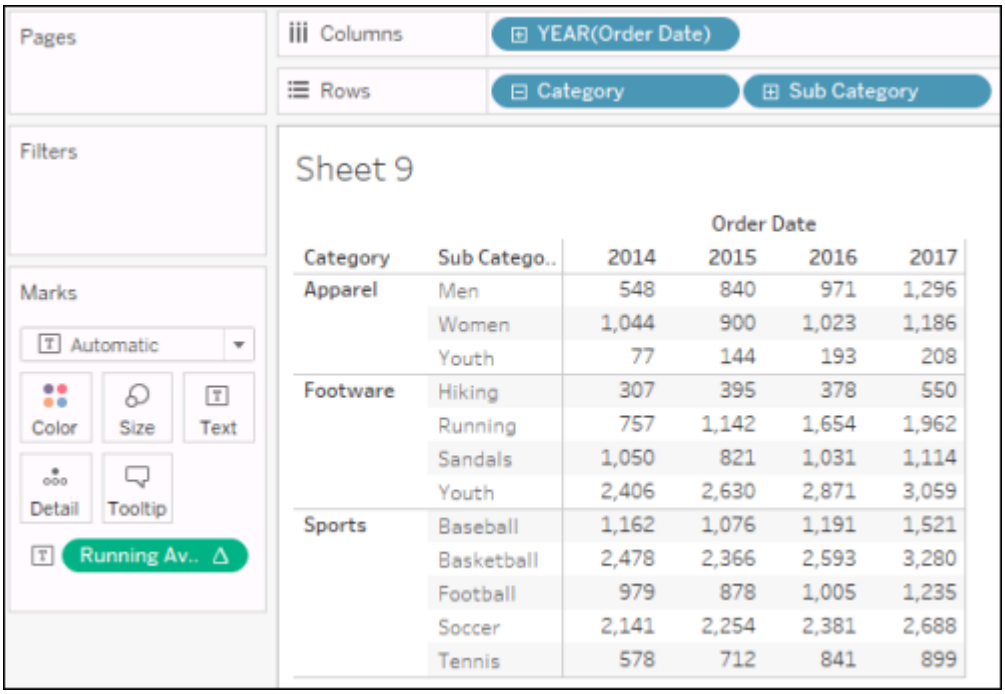
b) In the **Data** pane, under **Measures**, select and drag **Running Average** to the **Rows** shelf.

c) In the canvas, observe the running average values for profit and notice the **Sports** category has the highest average.




d) In the **Rows** shelf, select and drag **Running Average** to the **Label** box in the **Marks** card.

e) In the **Rows** shelf, in the **Category** pill, select the **Plus Sign** to expand **Category** to show the subcategory.

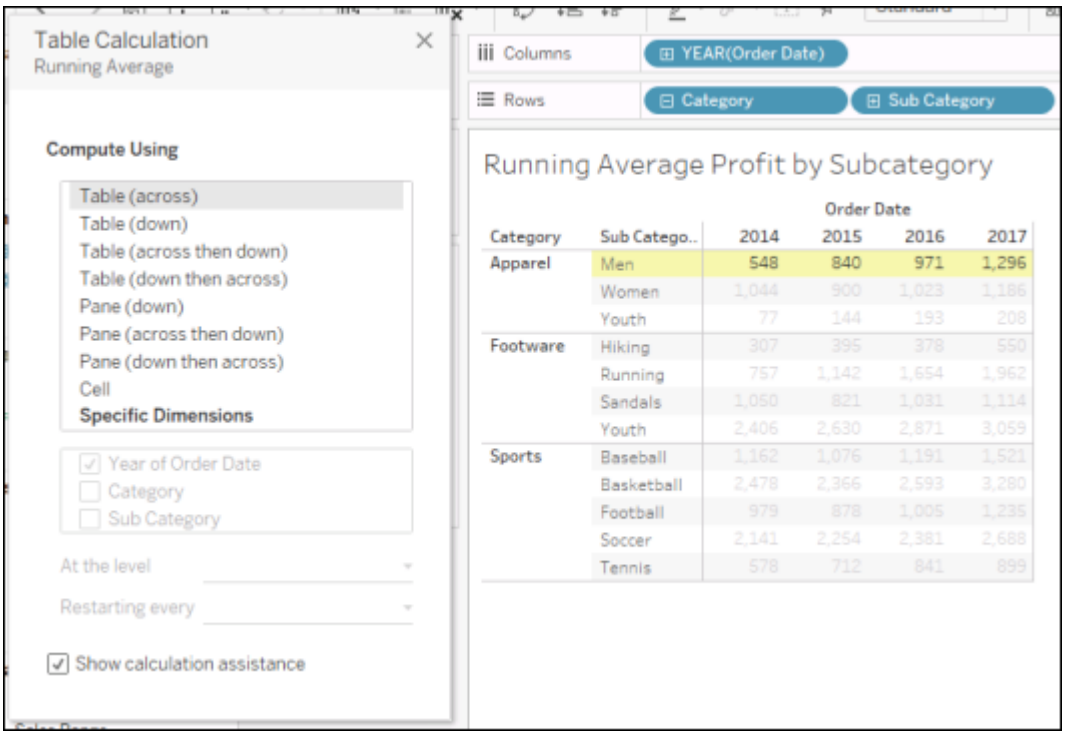



5. Rename the worksheet.

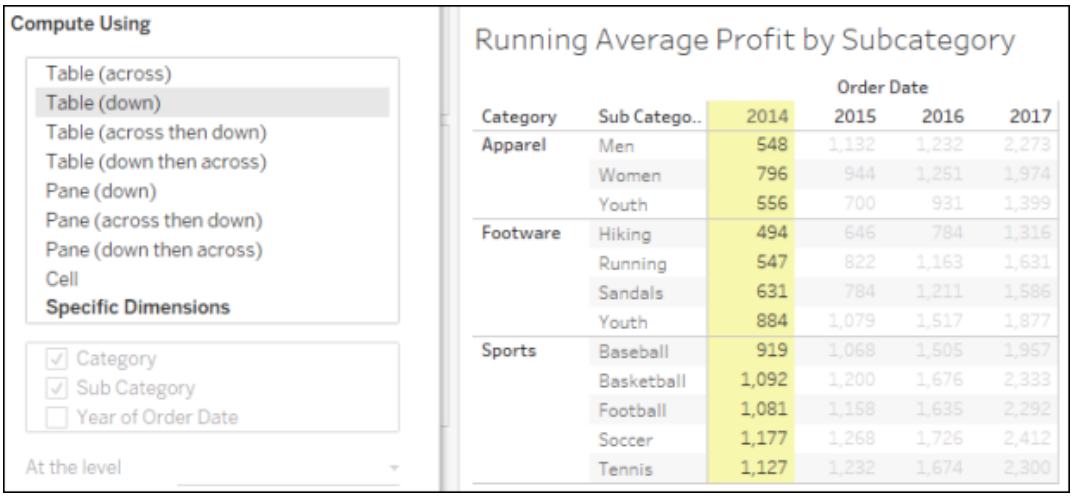
a) Right-click the **Sheet 9** tab and select **Rename**.
b) Type ***Running Average Profit by Subcategory*** and press **Enter**.

## 6. Change the **Compute Using** setting.

a) In the **Marks** card, right-click **Running Average**, and then select **Edit Table Calculation**.
b) Position the **Table Calculation** dialog box so it and the **Running Average Profit by Subcategory** canvas are both visible.
c) Observe that the current setting of **Table (across)** only goes from left to right for each subcategory when performing the running average. This would be useful if you want to see the running average for each subcategory across four years.

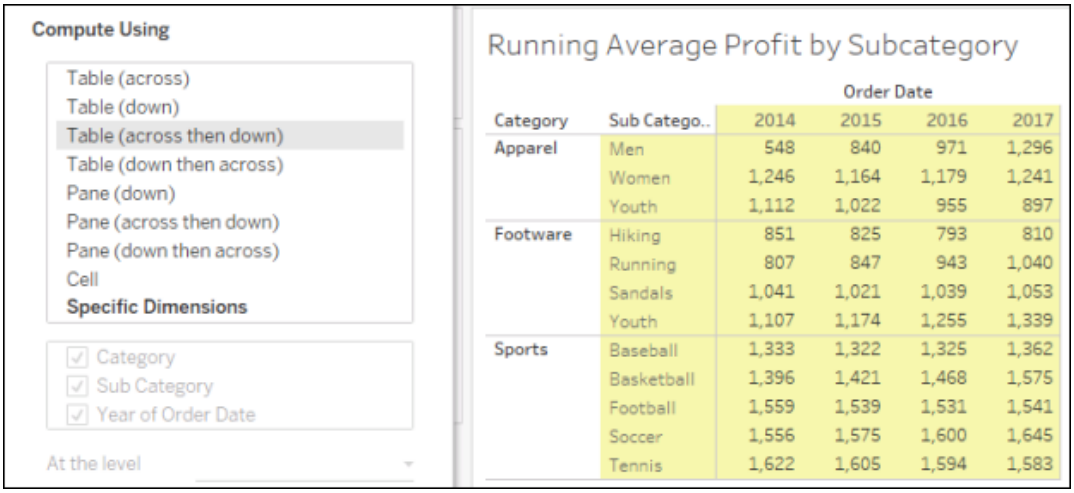



d) In the **Table Calculation** dialog box, select **Table (down)**.
e) Observe the current setting of **Table (down)** only goes from top to bottom for each year when performing the running average. This would be useful if you want to see the running average for each year across all of the subcategories.



f) In the **Table Calculation** dialog box, select **Table (across then down)**.

g) Observe the current setting of **Table (across then down)** goes from left to right for each subcategory and then also goes from top to bottom for each year when performing the running average. This would be useful if you want to see the running average for each subcategory across the four years and then continue averaging down to each subcategory.




h) Close the **Table Calculation** dialog box.

## 7. Save the workbook.

a) From the menu, select **File→Save**.

b) From the menu, select **File→Close**.

# Summary

In this lesson, you created calculated fields and in-shelf calculations to transform data to suit your analysis needs. You manipulated numeric, date, and string data using functions to find insights in the data. Finally, you analyzed data using table calculations and quick table calculations.

### What types of calculations and functions are you most likely to use in your environment?

### In what scenarios will you use table calculations?

**Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience.

# 6 Visualizing Data with Advanced Calculations

**Lesson Time: 1 hour**

## Lesson Introduction

Not only can calculations manipulate data and help in data analysis, calculations can also be used to affect the elements of visualizations. In this lesson, you will create groups and bins in Tableau® that are based on calculations, and perform data analysis with level of detail calculations.

## Lesson Objectives

In this lesson, you will:

- Create groups and bins with calculations.
- Analyze data with LOD expressions.

# TOPIC A

## Create Groups and Bins with Calculations

There are times when objects need to be organized for visualizations based on formulaic requirements. In this topic, you will create groups and bins based with calculations.

### Calculated Groups

As you know, you can create data groups in Tableau to combine related members into a single field. For example, if you are working with a list of products, you might want to create groups to combine all apparel items, all accessories, all men's shoes, all women's shoes, and so on. You can also create groups to correct flaws or inconsistencies in data, such as combining state names entered inconsistently.

Groups created visually in Tableau are static. Their membership does not change, and depending on the purpose of the group, may become out-of-date over time. To ensure group memberships remain up-to-date, you can create groups using calculations. The members of calculated groups update dynamically based on results of the calculation.

A calculated group is really just a calculated field you've chosen to use for grouping. It's simply a calculated field that groups things. If you wanted to see products that sold more than 20,000 units, you can create a calculated field with the following formula:

```
IF Sum([Unit Sales]) > 20000
THEN "Above 20K Units"
ELSE "Below 20K Units"
END
```

You can use this calculated field the way you would any calculated field in the Viz.

### Interactive Calculated Groups

You can add parameters to groups to give users interactive control over values in the calculation. For example, the following calculation groups products by those that have sold more than 20,000 units, and those that have sold less than 20,000 units.

If you want to give users the ability to choose the level of unit sales they want to set as a threshold, you can replace the number of units—the number 20,000—with a parameter. Simply create a parameter with a data type of **Float**. Set the current value to 20,000 and name it [Unit Sales Threshold]. You can then edit the calculated field to use the parameter:

```
IF Sum([Unit Sales]) > [Unit Sales Threshold]
THEN "Above Threshold"
ELSE "Below Threshold"
END
```

To make the parameter available to use, right-click and select **Show Parameter Control**. From that point on, users can interactively adjust the number of units required to be above the threshold.

### Bins

A *bin* is a way to segment a continuous range of numbers into discrete groups for analysis. Remember the dimensions are discrete, having a finite range of values, whereas measures are continuous, potentially having an unlimited set of values. At times, it's beneficial to segment measures for analysis by taking a discrete number of values from that continuous range to use for analysis. This is called creating a bin, or a binned dimension.

Bins are created from measures, and, once created, are listed as a dimension because they have a discrete set of values.

For example, if you wanted to do demographic analysis of your customers, you could segment your customers into age groups. Age is a continuous number because a person's age can be any number. You can create bins to segment customers by age, into ranges that match common demographics, such as 18-30, 31-45, 46-65, and 65 and over.

**Note:** Any dimension can be considered a bin.

Creating these bins would allow you to analyze and compare data based on the age range customers fall into.

### Limitations

Bins have the following limitations:

- Only data from relational data sources can be binned.
- Bins cannot be used in calculations.

## Create Bins

To create a bin, right-click the **Measures in the Data** pane and select **Create→Bins**.

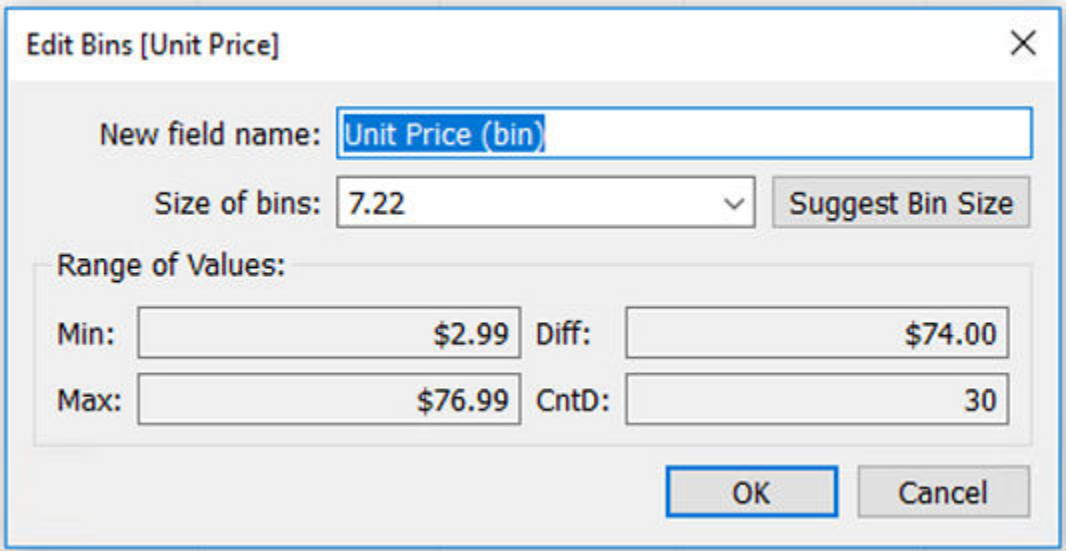



*Figure 6-1: Creating bins.*

Fill out the following information in the **Edit Bins** dialog box.

| *Configuration Object* | *Description* |
|---|---|
| **New field name** | Name of the bin. |
| **Size of bins** | Enter the size (number of times) each bin will contain. You can also click the **Suggest Bin Size** button to have Tableau suggest a bin size based on available data. |

| *Configuration Object* | *Description* |
|---|---|
| **Range of Values** | You cannot configure these items. Based on the field selected and the bin size, Tableau will tell you:<br>• **Min**: Minimum bin size.<br>• **Max**: Maximum bin size.<br>• **Diff**: Difference between the **Min** and **Max** size (**Max** – **Min**).<br>• **CntD**: The total number of bins. |

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Create Groups and Bins Based on Parameterized Calculations.**

# ACTIVITY 6-1

## Creating Groups and Bins Based on Calculations

### Data File

C:\095205Data\Refining Visualizations with Sets and Parameters\MyFootPrintSports_2018_6.twb

### Before You Begin

Tableau Desktop is open.

### Scenario

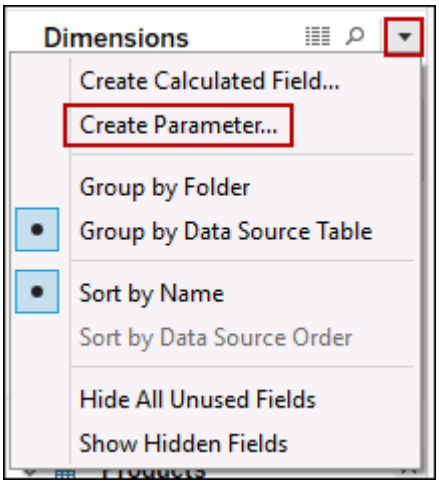
You want to alter your view for the top products by sales, only showing one category at a time. You want to be able to easily change from the different categories without having to edit the filter. You will use a parameter and a calculation to accomplish this. You also want to create bins for the number of days to ship orders so you can see how many orders take a certain number of days to ship.

---

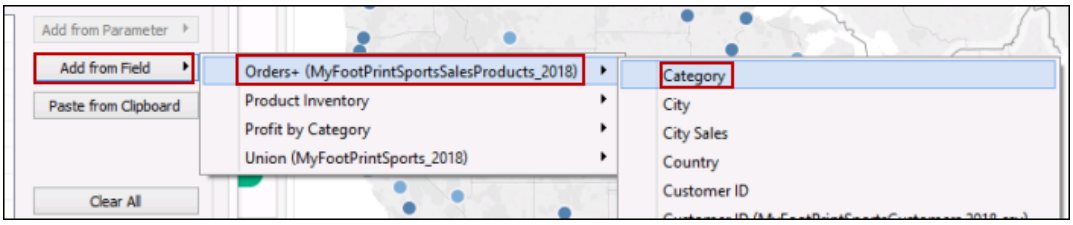
1. Open the worksheet for the lesson.
   a) From the menu, select **File→Open**.
   b) In the **Open** dialog box, browse to the **C:\095205Data\Visualize Data with Advanced Calculations** folder.
   c) Select **MyFootPrintSports_2018_6** and select **Open**.

2. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Visualize Data with Advanced Calculations** folder.
   c) In the **File name** box, type *MyCopy_MyFootPrintSports_2018_6* and select **Save**.

3. Create a parameter that has values for the categories.
   a) In the **Data** pane, to the right of **Dimensions**, select the drop-down arrow and select **Create Parameter**.



b) In the **Create Parameter** dialog box, in the **Name** box, type ***Categories***
c) In the **Data Type** drop-down list, select **String**.
d) For **Allowable values**, select **List**.
e) Select **Add from Field**, and then select **Orders+ (MyFootPrintSportsSalesProducts_2018)→Category**.



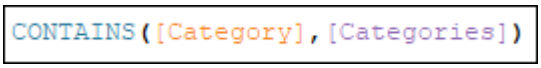

f) Select **OK**.

4. Create string calculation that uses parameter to select a category.

a) From the menu, select **Analysis→Create Calculated Field**.
b) In the dialog box, in the text box, type ***Product Categories***
c) In the code area, type `CONTAINS(`

d) In the **Data** pane, under **Dimensions→Products**, select and drag **Category** to the code area, at the end of the line.
e) In the code area, at the end of the line of code, type `,`

f) In the **Data** pane, under **Parameters**, select and drag **Categories** to the code area, at the end of the line.
g) In the code area, at the end of the line of code, type `)`




h) Select **OK**.

5. Use the string calculation to filter products displayed in a view.

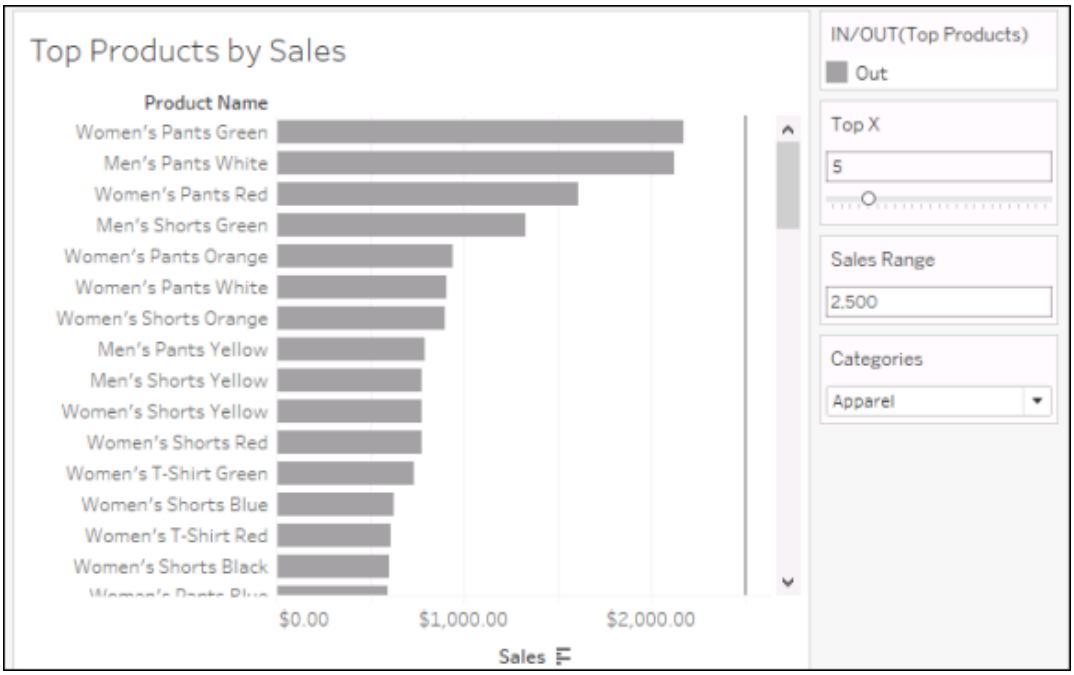
a) At the bottom of the screen, select the **Top Products by Sales** worksheet.

> **Note:** At this point, when selecting worksheets, you may need to navigate the worksheet tabs at the bottom of the Tableau window to find the worksheet you need. At the bottom of the screen, you can use the standard navigation buttons to scroll through the worksheets. You can also use the **Show Sheet Shorter** button to display all of the worksheets and then double-click the one you want.

b) In the **Data** pane, under **Dimensions**, select and drag **Product Categories** to the **Filters** card.
c) In the **Filter [Product Categories]** dialog box, in the list, check the **True** check box.
d) Select **OK**.

e) In the **Data** pane, under **Parameters**, right-click **Categories** and select **Show Parameter Control**.

f) In the canvas, observe the list of products and in the legend, observe the **Categories** card is currently set to **Apparel**.




g) In the legend, in the **Categories** card, in the drop-down list, select **Footware** and then **Sports** and observe the changes in products.

## 6. Copy and edit the **Days to Ship** calculated field to remove any aggregation.

a) In the **Data** pane, under **Measures**, right-click **Days to Ship** and select **Duplicate**.

b) Under **Measures**, right-click **Days to Ship (copy)** and select **Edit**.

c) In the **Days to Ship (copy)** dialog box, change the name to ***Days to Ship No Agg***

d) In the code area, at the beginning of the line of code, remove `INT(AVG(`.

e) In the code area, at the end of the line of code, remove `))`.

```
DATEDIFF('day',[Order Date],[Ship Date])
```

f) Select **OK**.

## 7. Create bins.

a) In the **Data** pane, under **Measures**, right-click **Days to Ship No Agg** and select **Create→Bins**.

b) In the **Edit Bins [Days to Ship No Agg]** dialog box, in the **Size of bins** box, type ***1***

You don't want to see partial days because you want to know how many full days it took for orders to ship.

c) Select **OK**.

## 8. Use bins in a view.

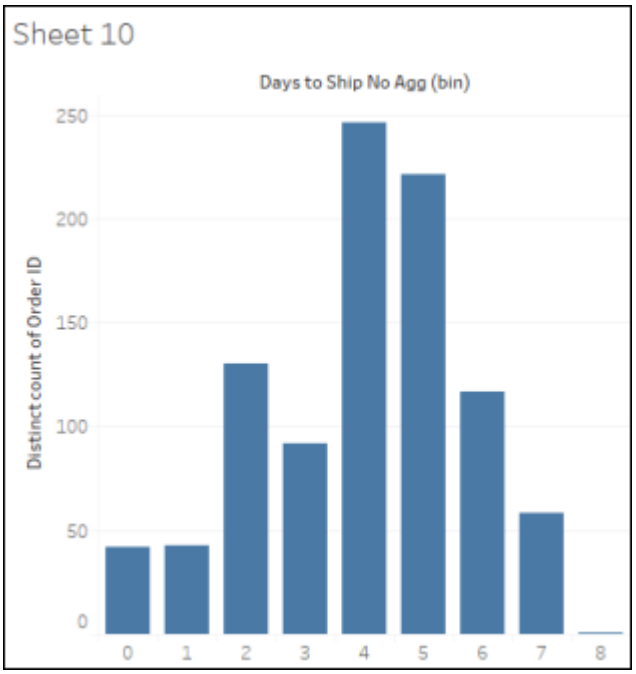
a) On the toolbar, select the **New Worksheet** button.

b) On the **Sheet 10** worksheet tab, in the **Data** pane, under **Dimensions**, select and drag **Days to Ship No Agg (bin)** to the **Columns** shelf.

c) In the **Data** pane, under **Dimensions→Orders**, select and drag **Order ID** to the **Rows** shelf.

d) In the **Rows** shelf, right-click **Order ID** and select **Measure→Count (distinct)**.

e) In the canvas, observe the bins that the number of days orders took to ship.




9. Rename the worksheet.

a) Right-click the **Sheet 10** tab and select **Rename**.

b) Type ***Number of Days to Ship Bins*** and press **Enter**.

# TOPIC B

# Analyze Data with LOD Expressions

There are times when the view in a worksheet may not show data at the right level of detail for analysis. In this topic, you will use level of detail (LOD) expressions to analyze data at different levels of detail.

## LOD Expressions

By default, Tableau computes aggregations at the view level. ***Level of Detail (LOD) expressions*** were introduced to allow data analysts to create calculations at a different level of detail than the view. They allow you to compute values either at the data source or view level with more or less detail.

> **Note:** Prior to LOD expressions, if you wanted to visualize a different aggregation, you needed to change the view or create a new one.

For example, if you've created a view that shows product sales for a shoe company, the view may show the [Category] of shoes, such as Men's and Women's, and the [Subcategory] for each, such as dress shoes, athletic shoes, and boots. The view may be aggregated at the [Category] level, showing totals for the broad categories of shoes, Men's and Women's. You may wish to analyze each subcategory so you can see totals for each type of shoe in each category. To do that you need to aggregate at a more granular level of detail, [Subcategory] by [Category]. You can use an LOD expression to show this.

An LOD is a calculated field and is created like any other calculated field, and it can be dimensions or measures:

```
{INCLUDE [Subcategory]: SUM([Sales)}
```

You then name the field and enter your LOD expression. You can then use the LOD expression in the Viz by dragging the calculated field to a shelf. You can set the type of aggregation to use by right-clicking the pill, or set it as part of the LOD expressions.

### Examples of LOD Expressions

To see example use cases of LOD expressions, see **https://www.tableau.com/about/blog/LOD-expressions**.

### Computation of Expressions in Tableau

Tableau computes expressions for unaggregated columns from that data source row-by-row for the underlying table. An example is a row-level expression. Expressions on aggregated data in source columns are computed as defined by the dimensions in the view. For more information, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_lod_overview.html**.

## Limitations of LOD Expressions

LOD expressions have the following limitations:

- Level of detail expressions that reference floating-point measures can behave sporadically in views that require a comparison of the values in the expression.
- Level of detail expressions are not shown on the **Data Source** page.
- Use the parameter name, not the value, when referencing a parameter in a dimensionality declaration.

- In data-blending scenarios, the linking field from the primary data source must be added to the view before you can use an LOD expression from the secondary data source.

## LOD Syntax

The syntax of an LOD expression is as follows. First, you tell Tableau the level of detail and the dimensions you wish to act upon to achieve that level of detail, then you enter the expression desired to perform the calculation to achieve the desired result. See the following step-by-step construction of a LOD expression.

LOD expressions must be surrounded by curly brackets {}:

```
{}
```

Next, you need an LOD keyword. This is **Fixed**, **Include**, or **Exclude**. The keyword does not need to be in all caps, but it's a good habit to get into:

```
{INCLUDE }
```

Next, you need to add the dimension you wish to act on. From the previous example, you want to act on the [Subcategory] field. You may include multiple fields in this portion of the expression separated by commas:

```
{INCLUDE [Subcategory]}
```

Type a colon to tell Tableau that you're done specifying the level of detail:

```
{INCLUDE [Subcategory]: }
```

Next, type the aggregate expression. This is the calculation you wish to perform to get the results you want in the view. You want to see [Sales] for each subcategory:

```
{INCLUDE [Subcategory]: SUM[(Sales)}
```

You must make sure the LOD expression is closed by the curly bracket. You can also specify aggregation as part of the LOD expression. For example, if you wanted to see the average sale by subcategory, you could wrap the entire expression in the aggregation statement:

```
AVG({INCLUDE [Subcategory]: SUM[(Sales)})
```

## LOD Keywords

LOD keywords are used to specify the level of detail for the calculation by telling Tableau which fields to include in the calculation. The following table explains the three keywords and their effect on the level of detail.

| *Keyword* | *Description* | *Yield* |
|---|---|---|
| **Fixed** | Performs the calculation using the dimension(s) specified in the level of detail portion of the expression and no other dimensions in the view. | Yields a measure if the field calculated is a measure. Yields a dimension if the field is a dimension. |
| **Include** | Performs the calculation using the dimension(s) specified in the level of detail portion of the expression *and all other dimensions in the view.* | Always yields a measure. |
| **Exclude** | Performs the calculation, ignoring the dimension(s) specified in the level of detail portion of the expression. Essentially, this performs the calculation using all dimensions in the view except those that are excluded. | Always yields a measure. |

Taking the previous example with a view that shows [Category] of shoes and [Subcategory] of shoes, if you wish to see aggregation at the level of detail of [Category] by [Subcategory], this LOD expression will achieve that result:

```
{INCLUDE [Subcategory]: SUM[(Sales)}
```

The expression achieves the desired result because of the INCLUDE keyword. Other fields in the view are included in the aggregation so the [Category] field is included with [Subcategory] to achieve the [Category] by [Subcategory] level of detail. If new fields are added to the view, they will also be included, which may change the calculation.

If you wanted to make sure the calculation is always delivered at the [Category] by [Subcategory] level of detail, you could create an expression using the fixed keyword:

```
{FIXED [Category],[Subcategory]: SUM[(Sales)}
```

## Level of Detail Aggregation

Since the LOD expression allows you to view data at a different level of aggregation than the view, the results of an LOD expression will likely be more granular or less granular than those in the view. The relationship between the level of detail in the view, and the level of detail in the LOD expression determines how Tableau will display the output from an LOD expression.

If the LOD expression is more granular than the view, then Tableau uses the aggregation from the LOD expression to display the results. If the LOD expression is less granular than the view, Tableau uses replication to show the results.

For example, if a shoe store created a sheet that has the [Segment] dimension in the view to show sales, the sheet would show sales for each of its three main market segments: women, men, and children. This view is considered aggregated to the level of detail of each [Segment]. If you wanted to see sales for the three main product types—dress shoes, boots, and athletic shoes—you could write the following LOD expression:

```
{FIXED [Segment],[Product Type]: Sum (Sale)}
```

This expression would result in a level of detail that is more granular than the view [Segment] by [Product Type]. Tableau would use the aggregation of the LOD expression to show the SUM of sales for each product type for each segment, as in the following figure.

Dimensions in the sheet:

- Segment

LOD Expression

{FIXED [Segment],[Product Type]: Sum (Sale)}

View is aggregated to [Segment] level.

| Segment | Sales |
|---|---|
| Women's | $ 732,131 |
| Men's | $ 706,730 |
| Children's | $ 626,162 |

RESULT: Sales is aggregated at a MORE granular level, [Segment] by [Product Type].

| Segment | Product Type | Sales |
|---|---|---|
| Women's | Dress shoes | $ 298,335 |
| | Boots | $ 198,415 |
| | Athletic shoes | $ 235,381 |
| Men's | Dress shoes | $ 321,412 |
| | Boots | $ 172,800 |
| | Athletic shoes | $ 212,518 |
| Children's | Dress shoes | $ 43,135 |
| | Boots | $ 135,215 |
| | Athletic shoes | $ 447,812 |

*Figure 6-2: Level of detail aggregation.*

## Replication

In the case where the same shoe store has a sheet at the level of detail of [Segment] by [Product Type], if you wished to see the SUM of sales for each [Segment] only, you could create the following LOD expression:

```
{FIXED [Segment]: Sum (Sale)}
```

Since the LOD expression is at a less granular level of detail than the view, replication will be used to display the results of the expression by showing the same value for each [Product Type] in the view.

Dimensions in the sheet:
- Segment
- Product Type

View is aggregated to [Segment] by [Product Type] level.

| Segment | Product Type | Sales |
|---|---|---|
| Women's | Dress shoes | $ 298,335 |
| | Boots | $ 198,415 |
| | Athletic shoes | $ 235,381 |
| Men's | Dress shoes | $ 321,412 |
| | Boots | $ 172,800 |
| | Athletic shoes | $ 212,518 |
| Children's | Dress shoes | $ 43,135 |
| | Boots | $ 135,215 |
| | Athletic shoes | $ 447,812 |

LOD Expression

{FIXED [Segment]: Sum (Sale)}

RESULT: Sales is aggregated at a LESS granular level at the [Segment] level and results are replicated.

| Segment | Product Type | Sales |
|---|---|---|
| Women's | Dress shoes | $ 732,131 |
| | Boots | $ 732,131 |
| | Athletic shoes | $ 732,131 |
| Men's | Dress shoes | $ 706,730 |
| | Boots | $ 706,730 |
| | Athletic shoes | $ 706,730 |
| Children's | Dress shoes | $ 626,162 |
| | Boots | $ 626,162 |
| | Athletic shoes | $ 626,162 |

*Figure 6-3: Replication.*

## Nested LOD Expressions

Nested LOD expressions can be useful when answering complex analysis questions. For example, if you wanted to see the average sales by product, you could use this LOD expression:

```
AVG ({FIXED [Product Name]: SUM([Sales)})
```

However, what if you wanted to see product sales for each product by state? You could then nest LOD expressions in the following way:

```
{FIXED [State]: AVG ({FIXED [Product Name]: SUM([Sales)})}
```

Nested calculations inherit their dimensionality from the outer calculation. The context of the inner calculation is defined by its parent, not the sheet. For example, if the inner calculation used the INCLUDE keywords as in the following LOD expression:

```
{FIXED [State]: AVG ({INCLUDE [Product Name]: SUM([Sales)})}
```

The calculation would yield the same result as this LOD expression:

```
{FIXED [State]: AVG ({FIXED [Product Name]: SUM([Sales)})}
```

This is because dimensionality is determined by the parent or outer expression.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Analyze Data with Advanced LOD Expressions.** |

# ACTIVITY 6-2

## Analyzing Data with Advanced LOD Expressions

### Before You Begin

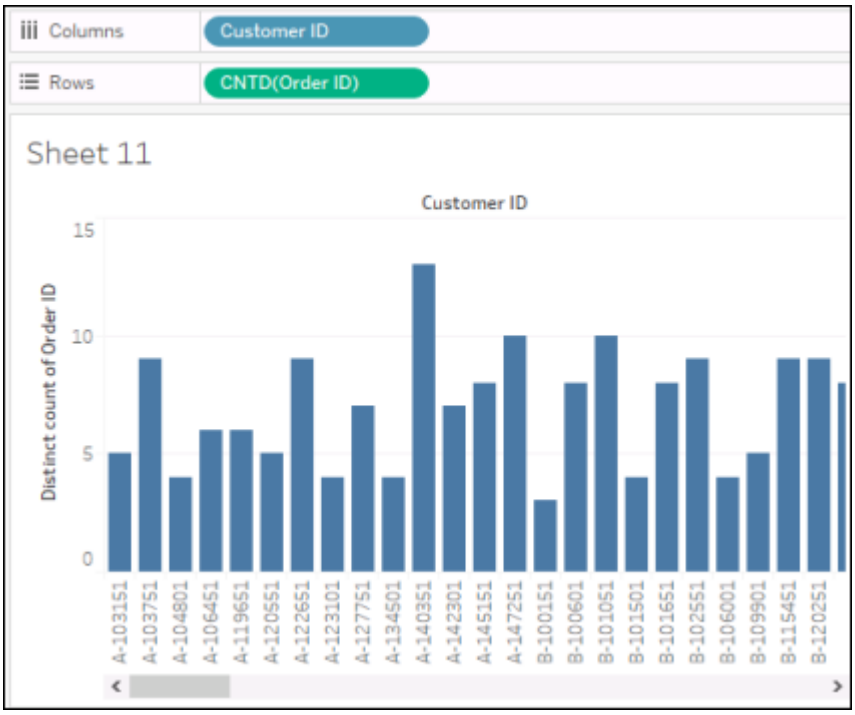
The **MyCopy_MyFootPrintSports_2018_6** workbook is open in Tableau Desktop.

### Scenario

You want to see the average number of orders across all customers. This will enable you to see how many customers make multiple purchases and the number of purchases on average. You will use LOD expressions to do this.

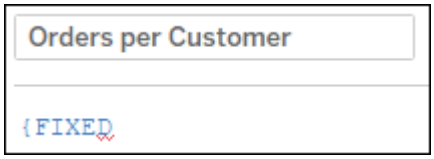
---

1. Create a view to show number of purchases made by each customer.
   a) On the toolbar, select the **New Worksheet** button.
   b) On the **Sheet 11** worksheet tab, in the **Data** pane, under **Dimensions→Orders**, select and drag **Customer ID** to the **Columns** shelf.
   c) In the **Data** pane, under **Dimensions→Orders**, select and drag **Order ID** to the **Rows** shelf.
   d) In the **Rows** shelf, right-click **Order ID** and select **Measure→Count (distinct)**.
   e) In the canvas, observe that you can see the number of orders each customer has purchased. However, you want to see how many customers have made a different number of purchases.



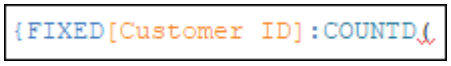

2. Create an LOD expression for the number of orders per customer.
   a) From the menu, select **Analysis→Create Calculated Field**.
   b) In the dialog box, in the text box, type *Orders per Customer*

c) In the code area, type `{FIXED`




d) In the **Data** pane, under **Dimensions→Orders**, select and drag **Customer ID** to the code area, at the end of the line.

e) In the code area, at the end of the line of code, type `:COUNTD(`




f) In the **Data** pane, under **Dimensions→Orders**, select and drag **Order ID** to the code area, at the end of the line.

g) In the code area, at the end of the line of code, type `)}`

h) Select **OK**.

i) In the **Data** pane, under **Measures**, select and drag **Orders per Customer** to **Dimensions**.

> **Note:** If **Orders per Customer** is left as a measure, then Tableau will treat it like a measure and apply aggregation like **Sum**, **Avg**, etc. to it. Making it a dimension allows it to act like a bin.

3. Edit the view for the number of purchases by customer.

a) On the toolbar, select the **Clear Sheet** button.

b) In the **Data** pane, under **Dimensions→Orders**, right-click and drag **Customer ID** to the **Rows** shelf.

> **Note:** Note that for this step and step **d**, you are using the right mouse button and not the left when selecting and dragging the **Customer ID** to the canvas. If you used the left mouse button. then simply remove the field and try again using the right mouse button.

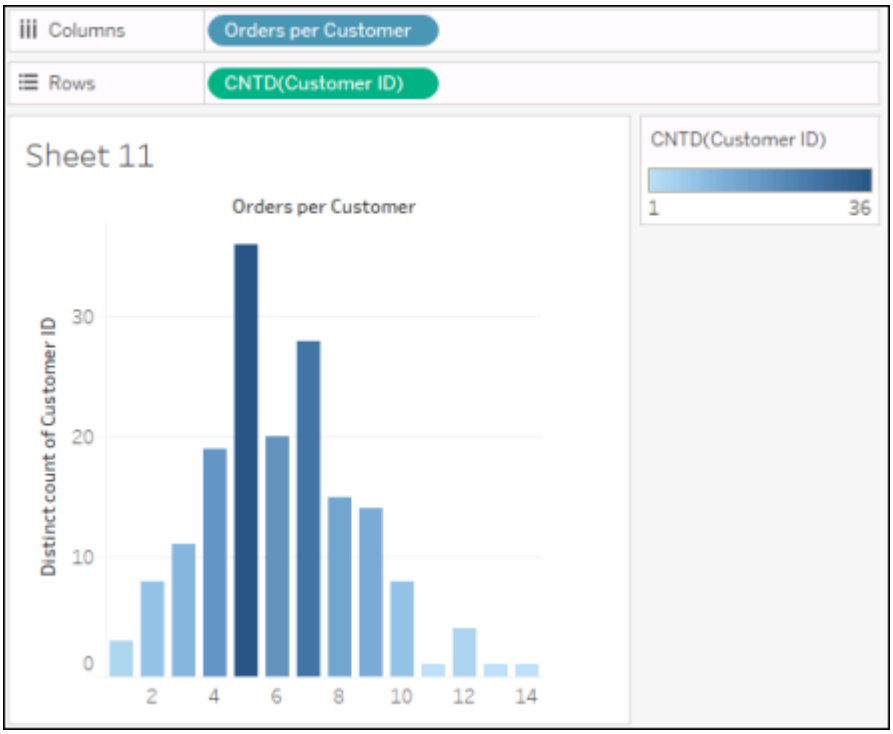
c) In the **Drop Field** dialog box, select **CNTD(Customer ID)** and select **OK**. This allows you to add the field with the **Count (Distinct)** measure.

d) In the **Data** pane, under **Dimensions→Orders**, right-click and drag **Customer ID** to the **Color** box in the **Marks** card.

e) In the **Drop Field** dialog box, select **CNTD(Customer ID)** and select **OK**.

f) In the **Data** pane, under **Dimensions**, select and drag **Orders per Customer** to the **Columns** shelf.

g) In the canvas, observe that you can see how many customers have made different numbers of orders with five orders being the most common for your customers.




4. Rename the worksheet.
   a) Right-click the **Sheet 11** tab and select **Rename**.
   b) Type ***Orders by Customer*** and press **Enter**.

5. Create a nested LOD expression that uses aggregation to calculate the average number of orders across all customers.
   a) In the **Data** pane, under **Dimensions**, right-click **Orders per Customer** and select **Duplicate**.
   b) Under **Dimensions**, right-click **Orders per Customer (copy)** and select **Edit**.
   c) In the **Orders per Customer (copy)** dialog box, change the name to **Avg Orders Across All Customers**
   d) In the code area, at the beginning of the line of code, type `{FIXED:AVG(` to add a second LOD expression that will average the results of the original LOD expression.

```
{FIXED:AVG({FIXED[Customer ID]:COUNTD([Order ID])
```

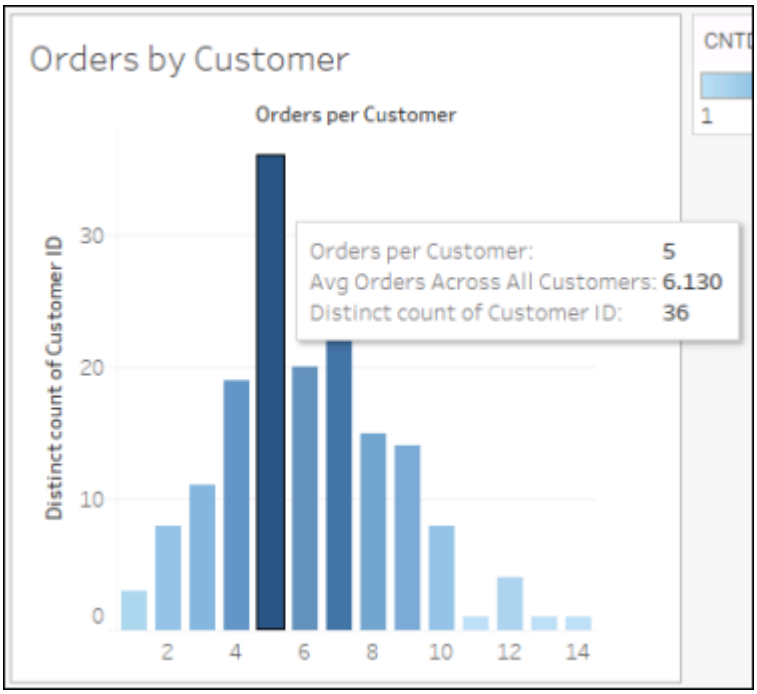
   e) In the code area, at the end of the line of code, type `)}`

```
{FIXED:AVG({FIXED[Customer ID]:COUNTD([Order ID])})}
```

   f) Select **OK**.

6. Add nested LOD expression to a view.
   a) In the **Data** pane, under **Dimensions**, select and drag **Avg Orders Across All Customers** to the **Tooltip** box in the **Marks** card.
   b) In the canvas, hover the mouse over one of the bars.

c) In the canvas, observe that the tooltip shows the average number of orders per customer across all customers.




7. Save the workbook.

a) From the menu, select **File→Save**.

b) From the menu, select **File→Close**.

# Summary

In this lesson, you created groups and bins based on calculations. You then used level of detail expressions to calculate fields to a different level of granularity than the view.

### Do you think you will use bin functionality in your environment? Why or why not?

### In which of your visualizations will you use LOD expressions?

**Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience.

# 7 Performing Statistical Analysis and Forecasting

**Lesson Time: 1 hour, 30 minutes**

## Lesson Introduction

While Tableau® allows any person who works with data to create visualizations and find insights, Tableau also supports many features that can be used for rigorous statistical analysis. In this lesson, you will perform statistical analysis and forecasting with Tableau.

## Lesson Objectives

In this lesson, you will:

- Perform statistical analysis of data using trend lines.
- Forecast data trends using Tableau forecasting options.

# TOPIC A

## Perform Statistical Analysis

Tableau's data connection and visualization capabilities make it an excellent platform for performing and visualizing statistical analysis. In this topic, you will perform statistical analysis of data in Tableau.

### Standard Deviation

One of the most common types of statistical analysis is finding the ***standard deviation***. Standard deviation measures how spread out a set of data is from a central line called the mean. Data points are typically clustered above and below the mean. With a low standard deviation, data is clustered closer to the mean and is considered more reliable. Data that is widely distributed around the mean denotes a high standard deviation and is considered less reliable.

A mean with a low standard deviation is more reliable because more data points are close to the mean. In essence, more data backs up the mean as a reliable measurement. You can use that measurement to make decisions with more confidence. The opposite is true of a mean with a high standard deviation.

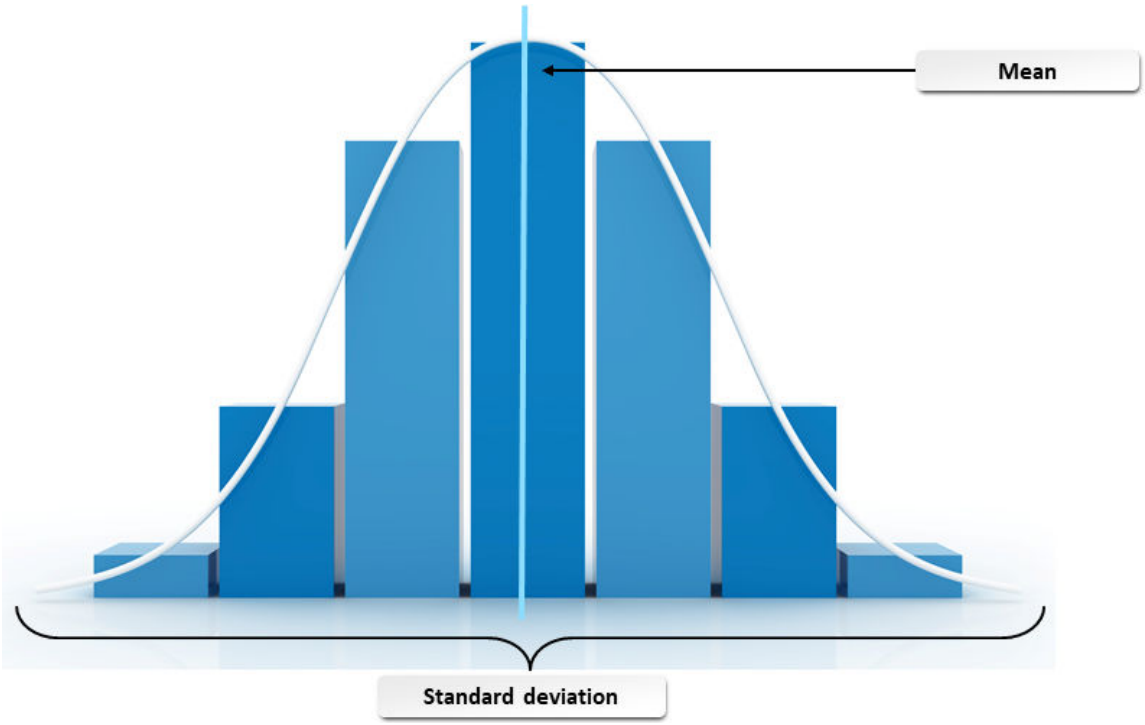



*Figure 7-1: Standard deviation.*

### Trend Line Analysis

Another common form of statistical analysis is trend line analysis. ***Trend lines*** show a statistical pattern or path over time. Most people are familiar with seeing business-related trend lines in graphics. When the line is going up from left to right over time, the business is doing well. Sales are up and the business is growing. If the trend is going down left to right, sales are falling and the business is failing. Trend lines show relationships between similar items, such as sales and customers. Analyzing these relationships can provide valuable insights.

For example, if you analyze product sales, you can see which products are growing in popularity with your customers, and which are decreasing. This will give you insights about which items to stock. You might also analyze customers to see the trend in purchases from small, medium, and large businesses. This information will help you choose the group to spend the most advertising dollars on.

You can add trend lines to your view in the **Analytics** pane. To add a trend line, both axes in the view must contain a field that can be interpreted as a number. You cannot add a trend line to a view that has product types that contain name strings as a dimension column and sales as rows. You could add a trend line if time measures are the columns and sales are the rows, because both can be interpreted numerically.

**Note:** For multidimensional data sources (cubes), date hierarchies contain strings rather than numbers, so trend lines are not available. Data sources with m/d/yy and d mmmm yyyy date formats do not allow trend lines.

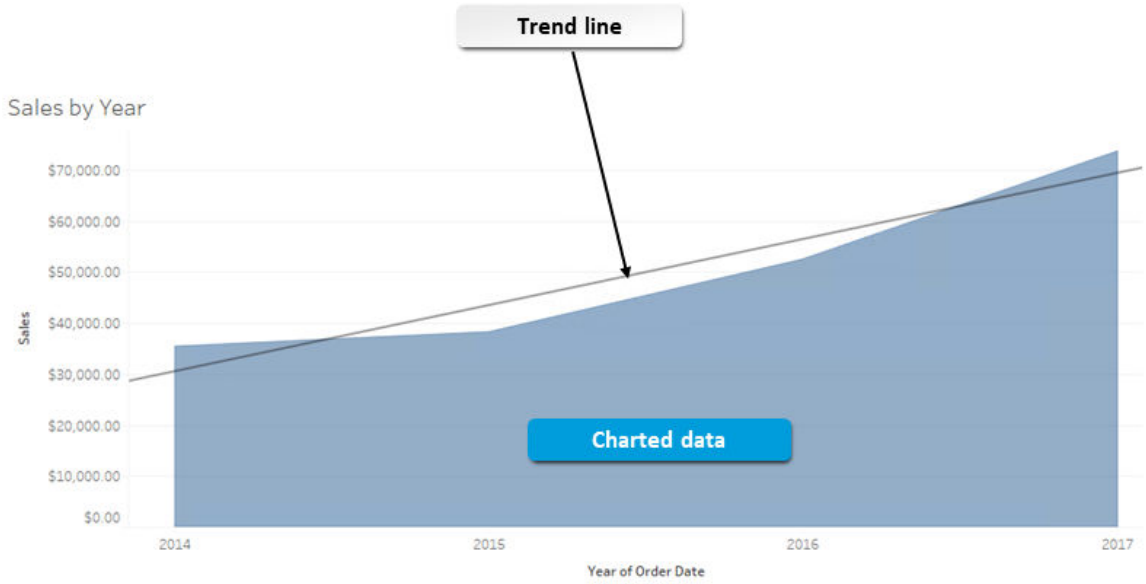



*Figure 7-2: Trend line analysis.*

## Trend Line Configuration

When adding a trend line, you can configure the following.

| *Configuration* | *Description* |
|---|---|
| **Trend line model** | Allows you to govern how the trend line is calculated.<br>Available models:<br>• Linear<br>• Logarithmic<br>• Exponential<br>• Power<br>• Polynomial |

| *Configuration* | *Description* |
|---|---|
| **Fields to use as factors** | Allows you to remove fields from the model when building trend lines that use multiple fields. For example, if you have a view showing a line graph with sales by region, leaving region as a factor will generate a different trend line for each region. Removing region as a factor will generate a single trend line, allowing you to evaluate each region versus the overall performance of all regions. |
| **Allows one trend line per color** | By default, trend lines are laid out per pane, and per color. If you drag another dimension into the view, you get another trend line. This box allows to you configure your trend line to ignore color-encoded data and use all data to build a single trend line. |
| **Confidence bands** | Show the upper 95% of confidence lines by default when trend lines are added. |
| **Force y-intercept to zero** | Allows you to force your trend line to start at zero. This is only available when both rows and columns have continuous fields. |
| **Show recalculated lines** | Allows you to configure if recalculated lines are shown when you select or highlight data in the visualization. |

## Coefficient of Determination: R-Squared

The coefficient of determination, commonly referred to as ***R-squared***, is another statistics term that helps validate the trend line. It tells you how well the model that makes up the trend line fits the data. The values can range from 0 to 1. Higher values are considered a better fit that validates the model. The value 1 would be a perfect fit; however, most statisticians are suspicious of very high values of 0.999 or better, feeling those may indicate problems with the trend model. Generally, values from 0.9 to 0.98 are an excellent fit.

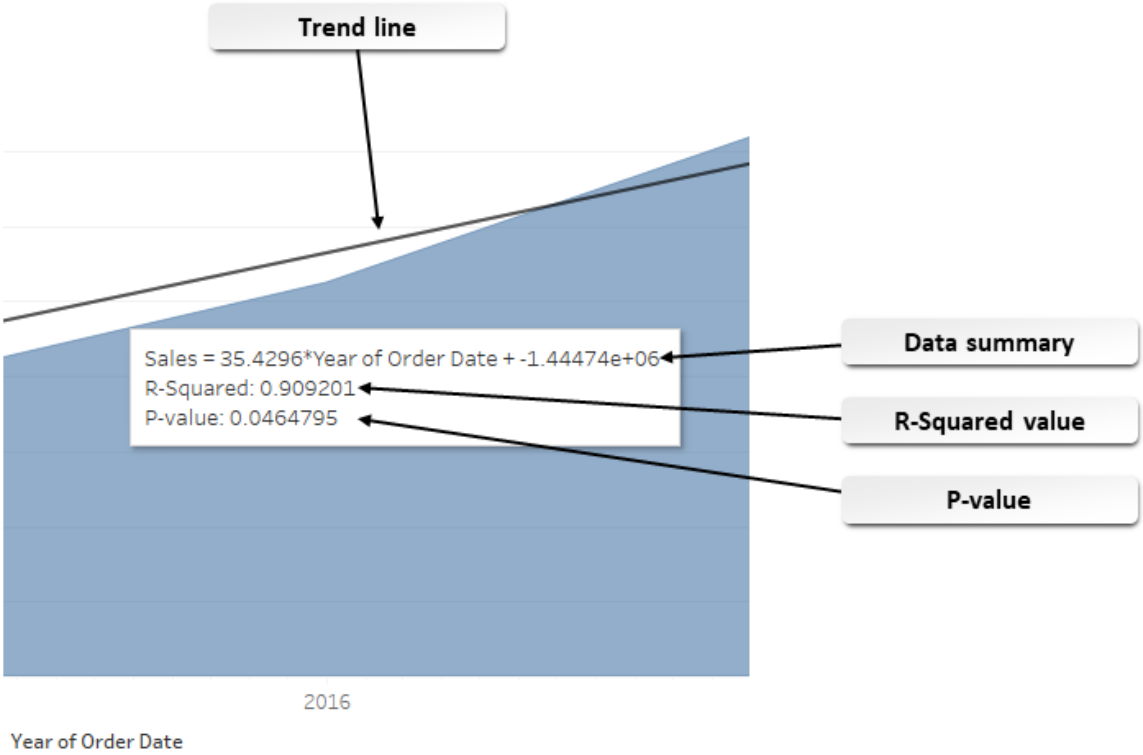



*Figure 7-3: Coefficient of determination: R-squared.*

## P-Value

When performing trend line analysis, it's important to validate that the trend line provides good insights based on the data you have. In other words, if the trend line is plotted accurately to reflect the data. Tableau allows you to analyze trend lines to validate they are showing statistically significant data. You can hover the mouse over a trend line and a tooltip will appear that shows this information.

One of these pieces of data is the ***probability value***, often called the P-value. The P-value is a number assigned to the significance of the trend line. All trend lines are made up of a data sample and the P-value helps confirm that, if you had more data, the same trend would remain true. To state it another way, the P-value helps confirm that, if you changed to a larger sample size, the same trend would be seen.

The P-value is a number between 0 and 1. There is a cutoff value assigned, usually 0.05. If the P-Value is less than the cutoff value, the results are considered significant. In other words, a large P-value greater than the cutoff on a scale of 0 to 1 can indicate support for the factors in the model, and indicates the trend line may be valid. The small P-value less than the cutoff indicates your model and trend line may be flawed.

## Residuals

P-value and R-squared should not be used alone to validate a trend line. Residual analysis should be used as well. In regression analysis, residuals are the difference between the measured value and value predicted by the model against the Y axis. If a model is good, plotted residuals should be randomly distributed around axis, near zero. If they are not, the model may not be the best fit for the data.

To create a residual plot you must first export the residual values for a trend line by selecting **Worksheet→Export→Data**. You can save the data in Microsoft® Access® format, and then connect to the data as a data source in Tableau and plot the data. This plot is a residual plot and can help validate the trend line.

## The R Project for Statistical Computing

The ***R Project*** is an open-source programming language and software environment for statistical computing and graphics. It is supported by the R Foundation for Statistical Computing. R Project and the R language are widely used by statisticians for data analysis. The ability to integrate R with Tableau opens up Tableau's robust visualization capability to statisticians who can use R and query R to perform statistical analysis from Tableau.

The R software environment consists of desktop software and a server component named R Serve, which are available for most operating systems. R supplies an environment that supports a wide array of linear and non-linear analysis and statistical tests, as well as graphing capabilities.

### Additional Information

For additional information on R, see **https://www.r-project.org/**.

## Tableau Integration with R

Both R and R Serve can integrate with Tableau. If you are familiar with R, know how to use the R language, and use R for statistical analysis, you can create functions that make calls to R and contain the R script. Tableau will then pass the script and data to R to perform the computation. R will return the result. This type of calculated field can be used like any other calculated field in Tableau.

To use R with Tableau, both Tableau Desktop and R software must be installed on the local machine. R Serve must be installed and initiated with R as well. Once all software is installed, you must create a connection to R Serve so that Tableau can pass function requests to R.

## Additional Information

For additional information on integrating R with Tableau, see **https://www.tableau.com/learn/whitepapers/using-r-and-tableau**.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Perform Statistical Analysis with Data.** |

# ACTIVITY 7-1

## Performing Statistical Analysis with Data

### Data File

C:\095205Data\Performing Statistical Analysis and Forecasting\MyFootPrintSports_2018_7.twb

### Before You Begin

Tableau Desktop is open.

### Scenario

You want to see the trends in your profit by subcategory view, so you will add a reference line to that view. You also want to use trend lines with your sales and profit by region and category view. You want the trend lines to be consistent within a single category.

---

1. Open the worksheet for the lesson.
   a) From the menu, select **File→Open**.
   b) In the **Open** dialog box, browse to the **C:\095205Data\Performing Statistical Analysis and Forecasting** folder.
   c) Select **MyFootPrintSports_2018_7** and select **Open**.

2. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Performing Statistical Analysis and Forecasting** folder.
   c) In the **File name** box, type *MyCopy_MyFootPrintSports_2018_7* and select **Save**.

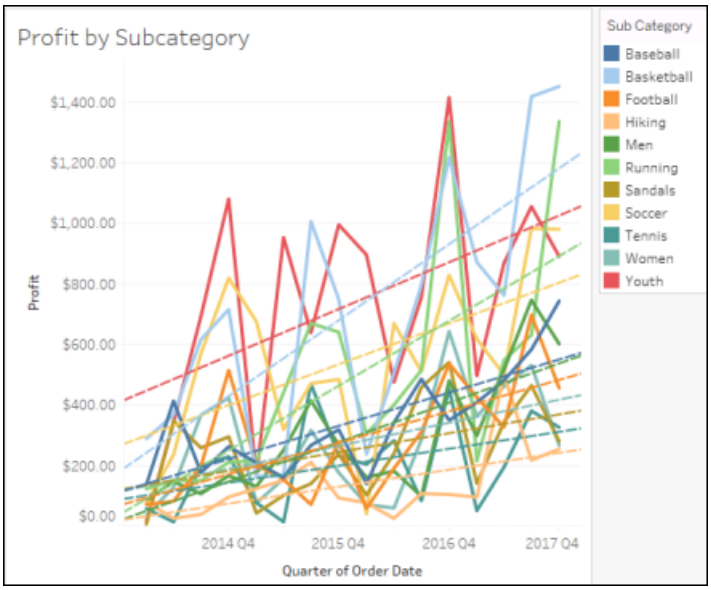
3. Add a trend line to a view.
   a) At the bottom of the screen, select the **Profit by Subcategory** worksheet.
   b) At the top of the **Data** pane, select the **Analytics** tab.
   c) In the **Analytics** pane, under **Model**, select and drag the trend line to the canvas and hover the mouse pointer over it without dropping it.

   | | |
   |---|---|
   | | **Note:** When dragging analytics items to the canvas, you need to hover the mouse pointer over the canvas and wait for the menu to appear. |

   d) From the menu, drop the trend line on **Linear**.

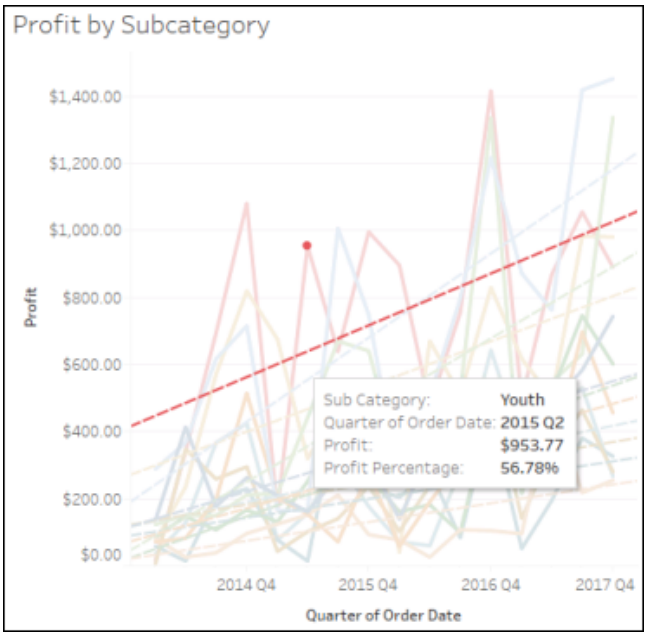
   | | |
   |---|---|
   | | **Note:** Tableau defaults to **Linear**, so if you dropped the trend line before dropping it on **Linear**, you can proceed normally. |

e) In the canvas, observe there is a trend line for each color in the view.




f) Select one of the trend lines to highlight it and make it easier to see.

g) Hover the mouse pointer over that trend line to display the R-squared and P-value of the line.




h) In the canvas, select the white space to deselect the trend line.

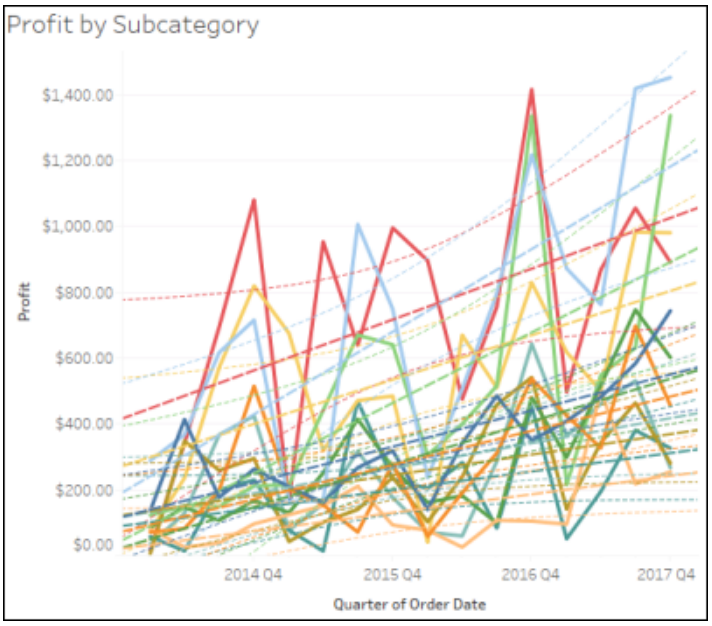
## 4. Edit a trend line.

a) From the menu, select **Analysis→Trend Lines→Edit Trend Lines**.

b) Position the **Trend Line Options** dialog box so you can view it and the canvas at the same time.

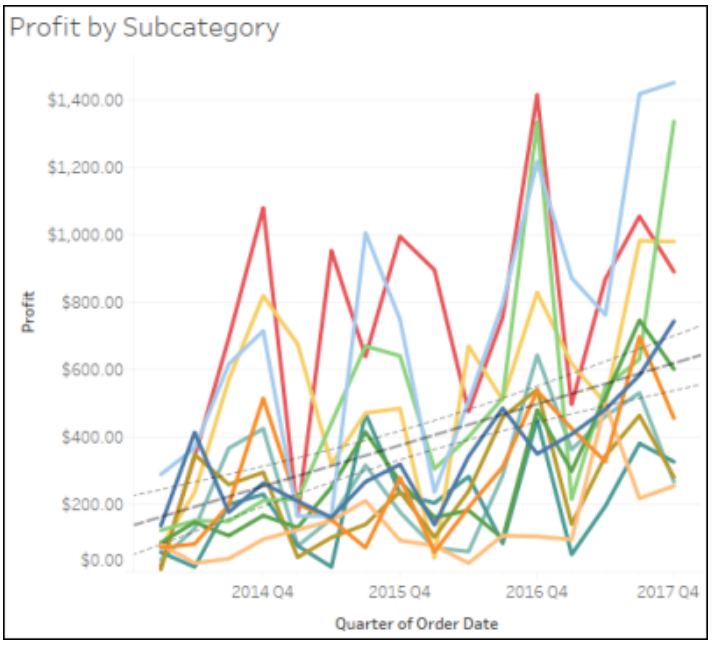
c) In the **Trend Line Options** dialog box, under **Options**, check the **Show Confidence Bands** check box.

d) In the canvas, observe there are confidence bands for each of the trend lines. With this many reference lines, it makes it difficult to see the information.




e) In the **Trend Line Options** dialog box, under **Options**, uncheck the **Allow a trend line per color** check box.

f) In the canvas, observe that there is now only one trend line with confidence bands showing the trend for the sum of all of the subcategories.




g) Select **OK**.

## 5. Use trend lines with another view.

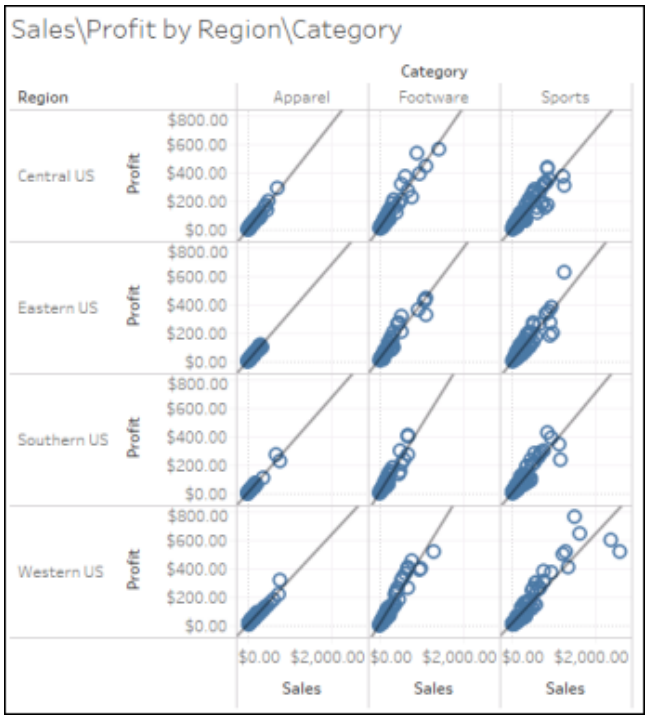
a) At the bottom of the screen, select the **Sales\Profit by Region\Category** worksheet.

b) In the **Analytics** pane, under **Model**, select and drag the trend line to the canvas and hover the mouse pointer over it without dropping it.

c) From the menu, drop the trend line on **Linear**.

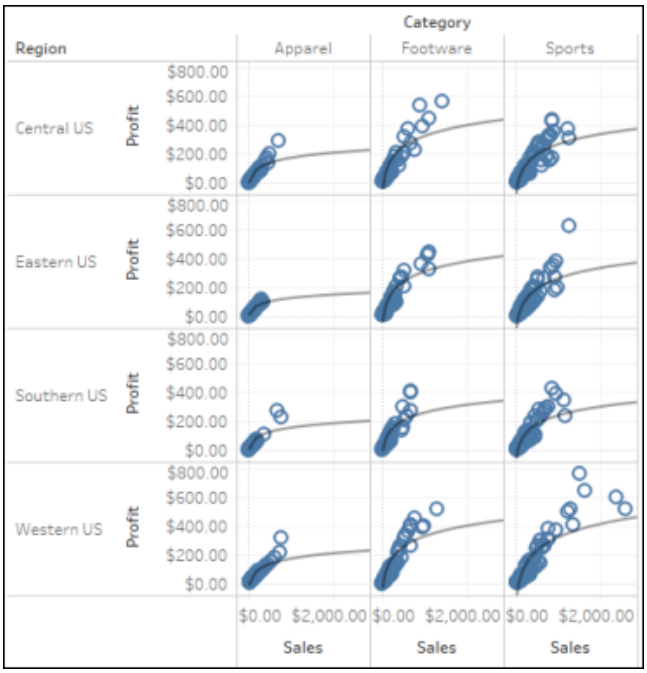
| | |
|---|---|
| | **Note:** Tableau defaults to **Linear**, so if you dropped the trend line before dropping it on **Linear**, you can proceed normally. |

d) In the canvas, observe there is a trend line for each pane in the view. Linear trend lines show a relationship between the variables that increases or decreases at a steady rate. They are best used when the trend of the data resembles a linear pattern.



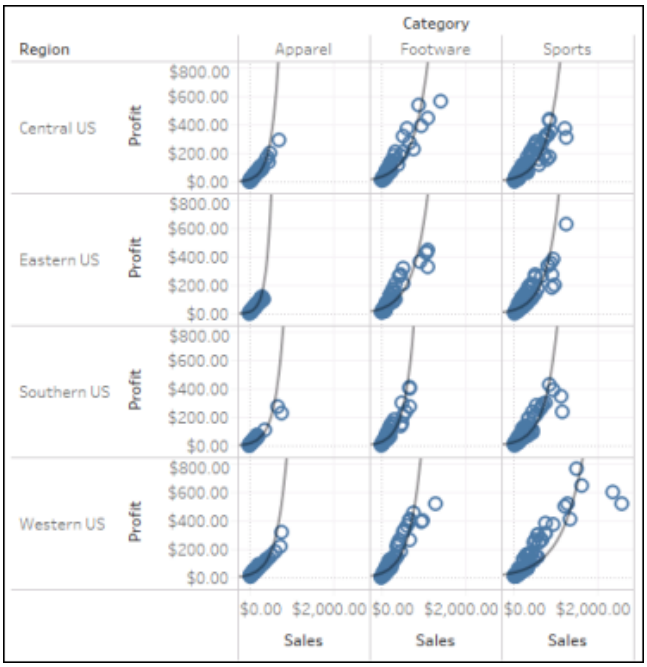

6. Explore different model types.
   a) From the menu, select **Analysis→Trend Lines→Edit Trend Lines**.
   b) Position the **Trend Line Options** dialog box so you can view it and the canvas at the same time.
   c) In the **Trend Line Options** dialog box, under **Model Type**, select **Logarithmic**.

d) In the canvas, observe the logarithmic trend lines, which are curved lines best used when the rate of change between variables increases or decreases quickly, but gradually levels out.



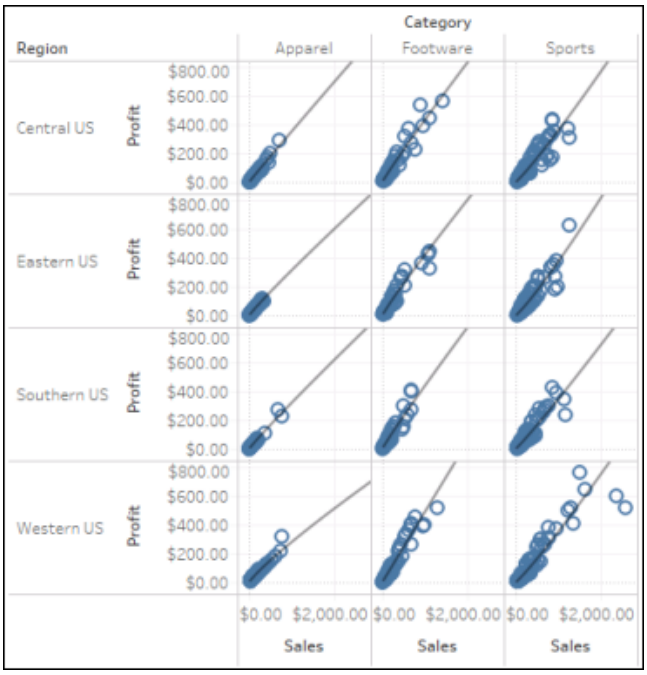

e) In the **Trend Line Options** dialog box, under **Model Type**, select **Exponential**.

f) In the canvas, observe the exponential trend lines, which are also curved lines that are best used when the rate of change between variables rises at increasingly higher rates.



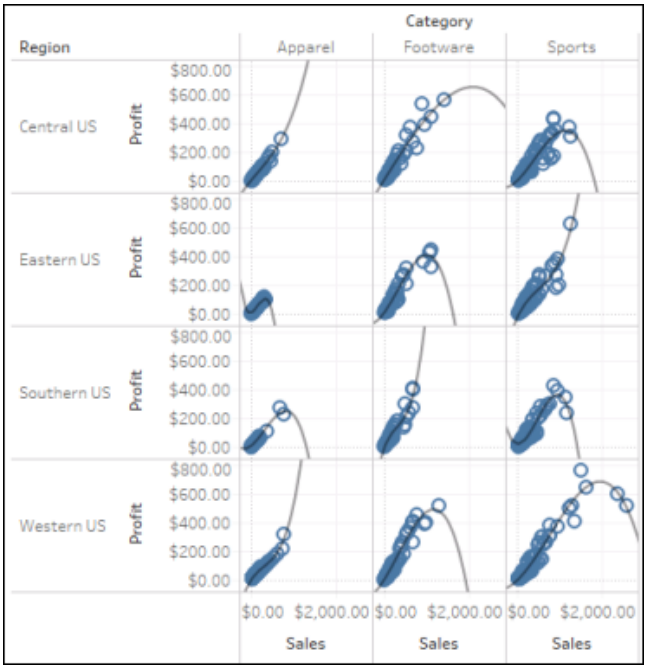

g) In the **Trend Line Options** dialog box, under **Model Type**, select **Power**.

h) In the canvas, observe the power trend lines, which are curved lines best used when the dependent variable increases at a specific rate.




i) In the **Trend Line Options** dialog box, under **Model Type**, select **Polynomial, Degree**.

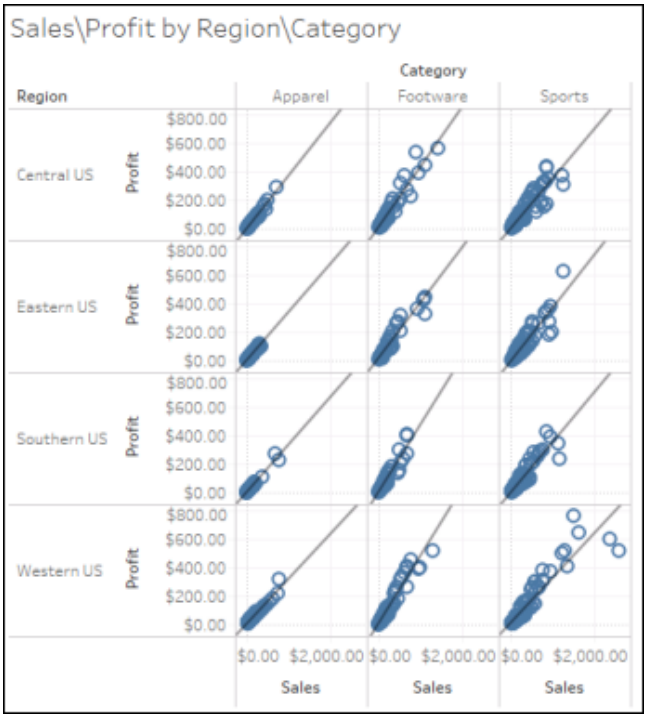
j) In the canvas, observe polynomial trend lines, which are curved lines best used when dealing with variables that have a fluctuating relationship. They have a value of 2 to 8 that can be defined by the user to make them assume different shapes to fit the relationship.




## 7. Change fields used as factors.

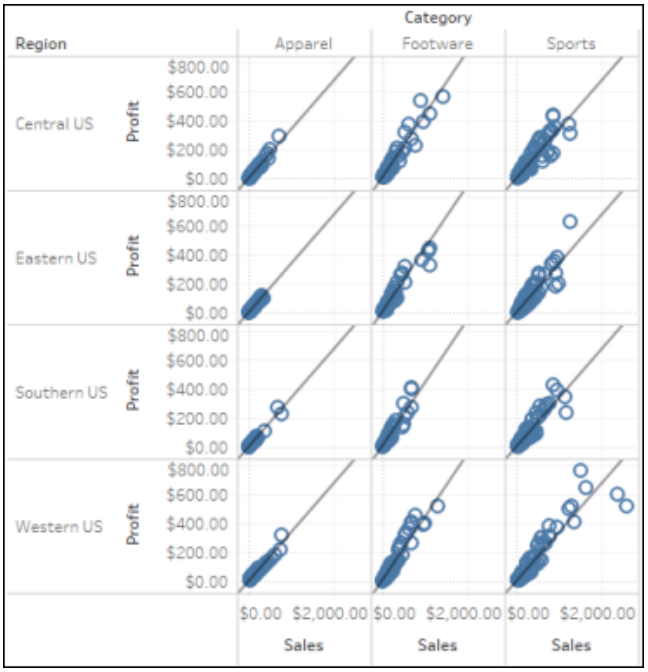
a) In the **Trend Line Options** dialog box, under **Model Type**, select **Linear**.

b) In the canvas, observe that each pane has its own model and trend line based on the category and region. You want the model to be the same across all regions so you can compare the profit numbers in relation to the same trend line.




c) In the **Trend Line Options** dialog box, under **Options**, uncheck the **Region** check box.

d) In the canvas, observe the trend lines for each category are the same across each of the regions.




e) Select **OK**.

# TOPIC B

## Forecast Data Trends

One of the most common forms of statistical analysis is creating forecasts so people and businesses can anticipate their needs as their businesses grow. In this topic, you will forecast data trends in Tableau.

### Forecasting

Forecasts use historical data and statistical models to make predictions of what will happen in the future. For example, if you have the past three years' worth of sales data, Tableau can create a forecast that predicts the coming year's sales. You can create a forecast from the **Analytics** pane by dragging the **Forecast** object into the view. When you create a forecast, by default, you will see the historical data and the forecast in separate panes in the view. You can adjust and customize the forecast view in many different ways.

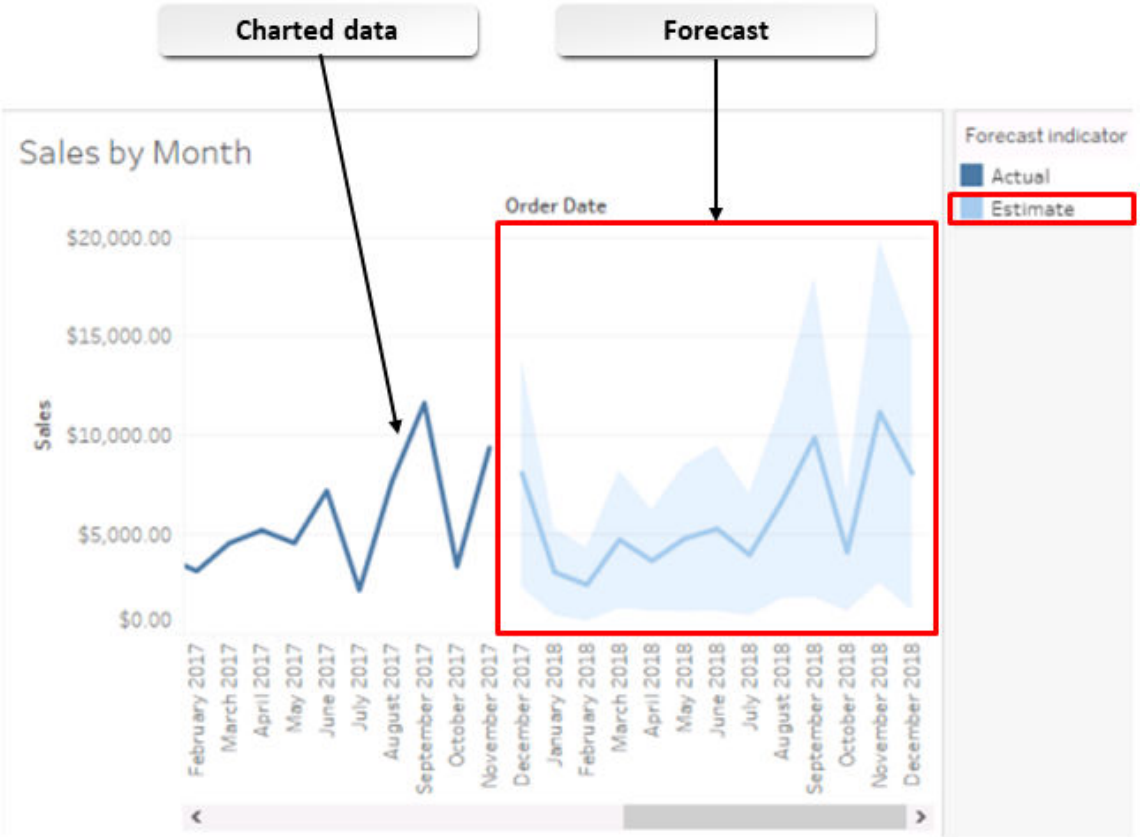



*Figure 7-4: Forecasting.*

Tableau uses an exponential smoothing model to create forecasts. Those models analyze past values using weighted averages to forecast new values and plot them over time for the length of the forecast. Tableau chooses the specific model automatically based on the data in the view. Tableau also tests the data in the view for seasonality, looking for seasonal changes based on the historical data. For example, some businesses, such as farms, have seasonal income and expenses. If Tableau detects that in the data, it will choose a forecasting model that will incorporate seasonality into the forecast.

To create a forecast, you must have at least one date or a dimension that contains integer values, and one measure with five data points. If you're forecasting seasonal data, you must have a minimum of two seasons' worth of data. Because of the models used in forecasting, forecasts cannot be created from ***multidimensional data sources***, such as cubes, views containing table calculations, totals or subtotals, or disaggregated measures or dimensions. Any pill that has been used to create a forecast will have an **Arrow** icon pointed in the upper-right direction.

## Prediction Interval

Forecasts also show ***prediction intervals***. For a line forecast, this is a shaded area around the line that shows the area where most of the actual values are likely to fall as forecast data is replaced by real data. If the forecast is a plot with points, whiskers will extend from plotted points to the prediction intervals.

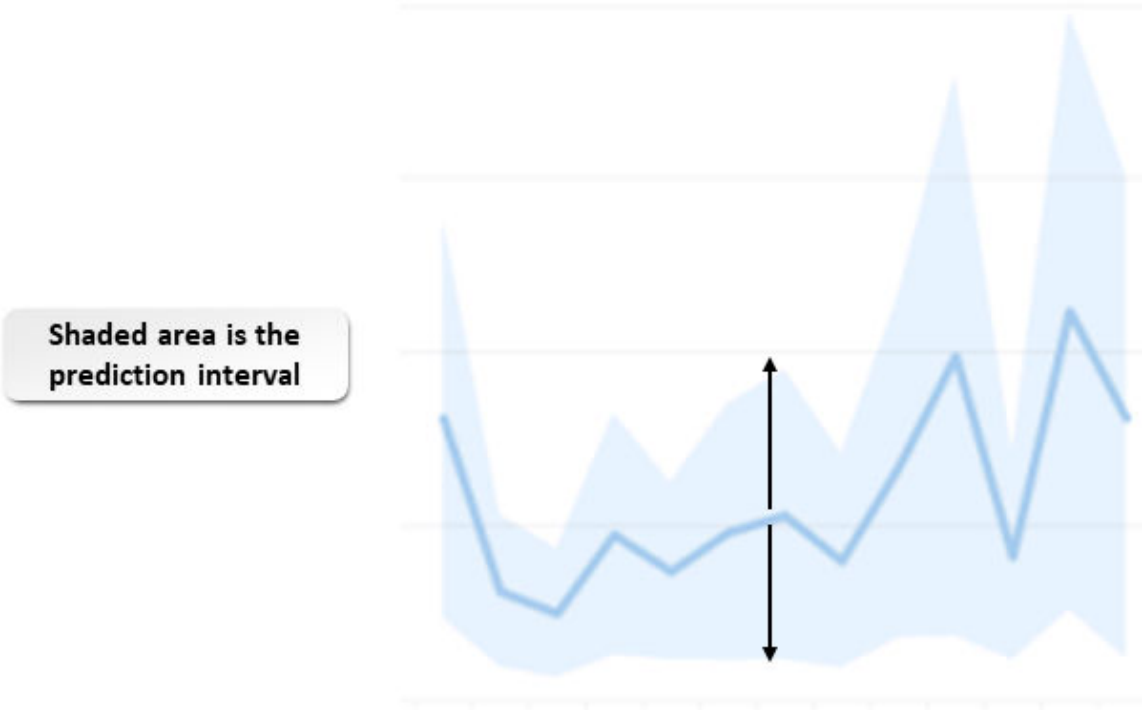



*Figure 7-5: Prediction interval.*

The default prediction interval is 95%, meaning that, based on the prediction model, 95% of actual values should fall near the forecast in the shaded area or within the scope of the whiskers, and only 5% of data will fall outside of the prediction interval. You can adjust the prediction interval in the **Forecast Options** dialog box. The higher the prediction interval, the more precise your confidence is, and the wider the band—or whiskers—will be.

## Forecast Options

You can adjust forecast options by right-clicking the pill used to create the forecast and selecting **Forecast Options**. You can configure the following settings in the **Forecast Options** dialog box.

| *Configuration Object* | *Description* |
|---|---|
| **Forecast length** | Configure how long you wish the forecast to extend into the future. You can specify a number and select **Years**, **Quarters**, **Months**, **Weeks**, **Days**, **Hours**, **Minutes**, or **Seconds**. The default is **Automatic** (next 12 months). |

| *Configuration Object* | *Description* |
|---|---|
| **Source data** | Select how Tableau will aggregate the data used for the forecast. You can select from the same time intervals available in **Forecast length**. The default is **Automatic** (months). |
| **Ignore** | It's often beneficial to ignore a portion of data that is incomplete. For example, if you're creating a monthly forecast in the middle of the month, you should exclude the last month of data because it is incomplete. By default, Tableau ignores the last unit of aggregation. |
| **Fill in missing values with zeros** | Check this box to fill any missing (null) values with zeros if your data is incomplete or has been joined or unioned and has null values. |
| **Forecast model** | You can select from automatic, automatic with seasonality (if you have seasonal data), and custom. If you select custom, you can configure both trend and season data by selecting from **None**; **Additive**, which sums contributions of model components; and **Multiplicative**, which multiplies contributions of model components. |
| **Show prediction intervals** | Check this box to show the prediction intervals. You can set the confidence level as well. The default confidence level is 95%. |
| **Description** | This area has a text description of the forecast. |

## Describe Forecast

If users right-click the pane containing a forecast, they can select **Forecast→Describe Forecast** to see a summary of what makes up the forecast and the models used.

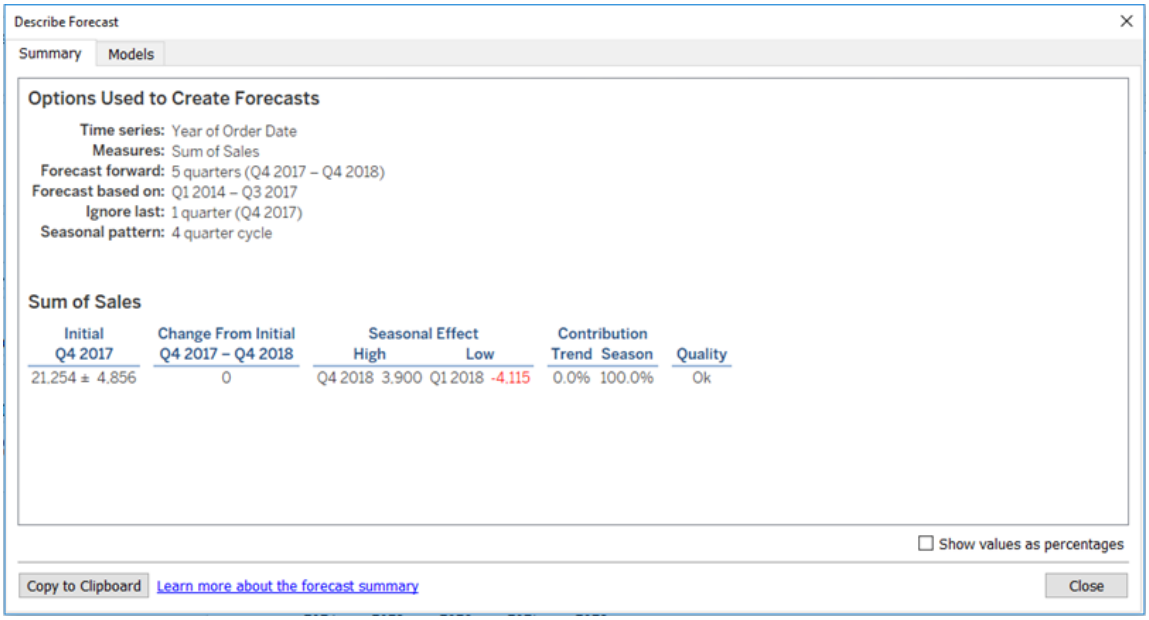



*Figure 7-6: Describe forecast.*

The dialog box has two tabs.

| Tab | Description |
|---|---|
| **Summary** | This tab shows the series and measure used in the forecast, as well as the length of the prediction. It also shows information about the data the forecast is based on, and any aggregation periods ignored and seasonality patterns found. It may also have additional information about the measure used. |
| **Model** | This tab describes the quality metrics and smoothing coefficients used in the model. |

**Note:** The **Describe Forecast** dialog box is informational only, as there is no way to edit or adjust the forecast configuration from it.

## Forecast Results

You can adjust the way in which the forecast is displayed by adjusting the forecast field results. To do so, right-click the pill used to create the forecast and select **Forecast result**.

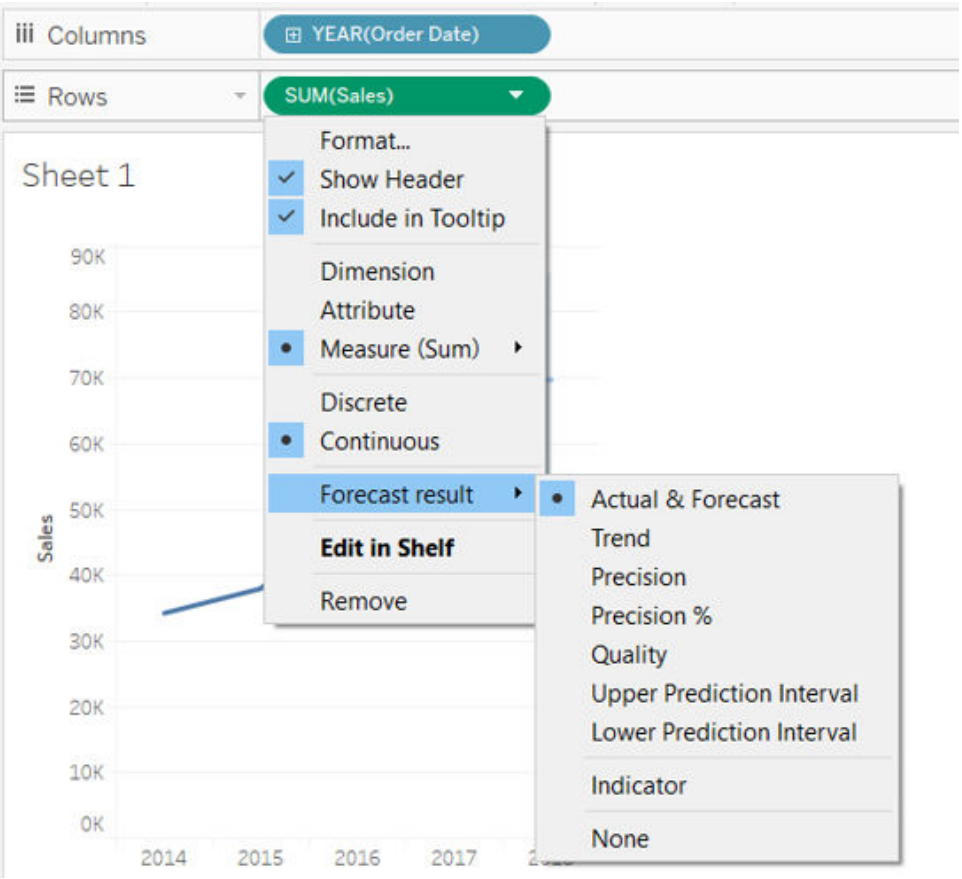



*Figure 7-7: Forecast results.*

You can then select one of the following options.

| Result | Description |
|---|---|
| **Actual & Forecast** | This is the default that shows data and the forecast. |
| **Trend** | Shows the forecast with seasonal components removed (if any). |

| *Result* | *Description* |
|---|---|
| **Precision** | Shows the prediction interval distance based on the confidence level set. |
| **Precision %** | Shows the prediction interval distance as a percentage of the forecast value. |
| **Quality** | Assigns the numbers on a scale of 0 to 100 (larger is better). Uses the mean absolute scaled error (MASE) scale of the forecast. This is an error weighting algorithm designed to give higher scores to better data. |
| **Upper Prediction Interval** | Shows a 90, 95, or 99 confidence level above the forecast value based on the set prediction interval. |
| **Lower Prediction Interval** | Shows a 90, 95, or 99 confidence level below the forecast value based on the set prediction interval. |
| **Indicator** | Shows **Actual** for rows in the worksheet when forecasting was inactive and **Estimate** for rows added when forecasting was activated. |
| **None** | Does not show forecast data. |

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Forecast Data Trends.**

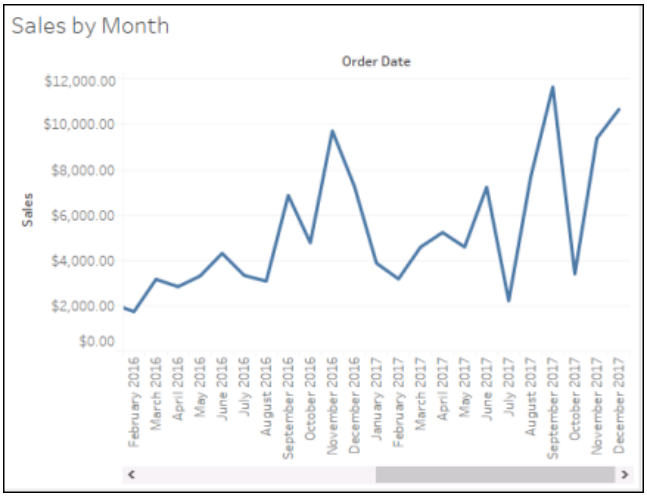
# ACTIVITY 7-2
## Forecasting Data Trends

### Before You Begin
The **MyCopy_MyFootPrintSports_2018_7** workbook is open in Tableau Desktop.

### Scenario
It is 2017 and you want to be able to plan for future sales. You decide to see how Tableau can help. You want to get an idea of what to expect for sales in the next six months, and will use the forecast feature to forecast those sales based on your previous sales data. You want to be as prepared as possible, so you will use the highest accuracy setting for the forecast.

---

1. Review the sales by month view.
   a) At the bottom of the screen, select the **Sales by Month** worksheet.
   b) In the canvas, scroll through the months and observe that this is a simple view that shows sales per month over four years.

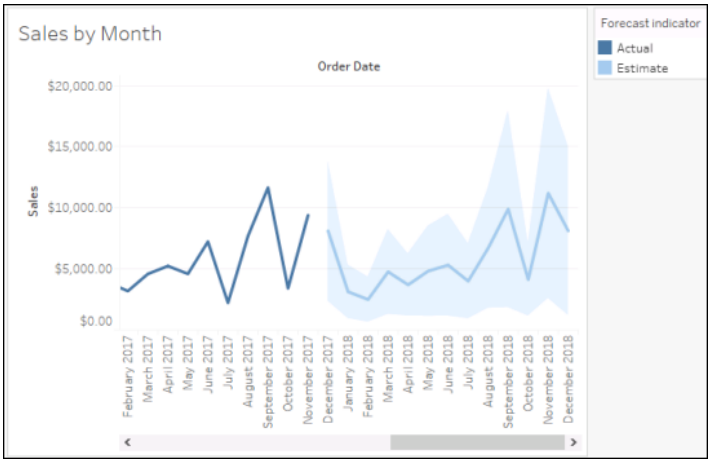



2. Add forecasting to the view.
   a) If the **Analytics** tab is not displayed, then at the top of the **Data** pane, select the **Analytics** tab.
   b) In the **Analytics** pane, under **Model**, select and drag **Forecast** to the canvas and hover the mouse pointer over it without dropping it.

| | |
|---|---|
| | **Note:** When dragging analytics items to the canvas, you need to hover the mouse pointer over the canvas and wait for the menu to appear. |

   c) From the menu, drop the **Forecast** on **Forecast**.

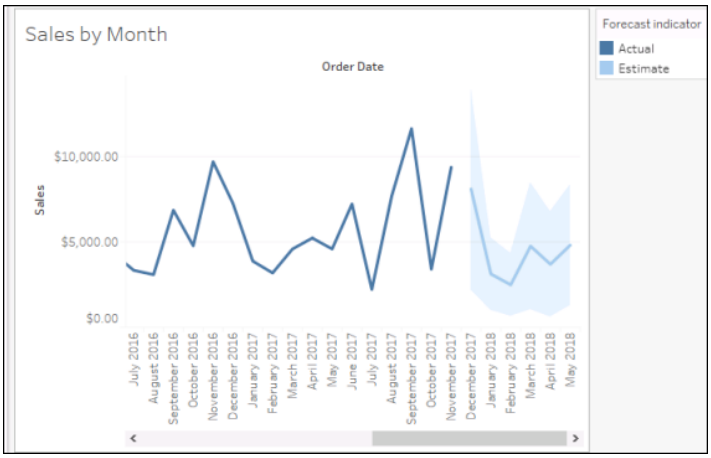
d) In the canvas, you may need to scroll to the right to observe that the forecast for sales in 2018 has been added to the view. The solid blue line is the forecasted data and the lighter shaded area is the prediction interval. The prediction interval is the range of variance for the data in the forecast.




e) Note the gap between the actual sales figures and the forecasted figures. Tableau ignores the last unit of aggregation—in this case a month—in case there isn't a full portion of data for that month.

3. Configure the forecast length.

a) From the menu, select **Analysis→Forecast→Forecast Options**.
b) Position the **Forecast Options** dialog box so you can view it and the canvas at the same time.
c) In the **Forecast Options** dialog box, under **Forecast Length**, select **Exactly**.
d) In the box, type ***6***
e) In the **Years** drop-down list, select **Months**.
f) In the canvas, observe that the forecast is now for just the next six months.

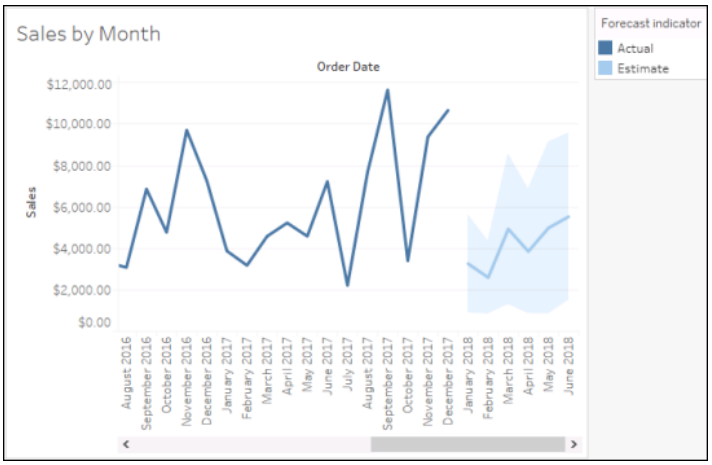



**Note:** If you cannot see the change in the view, close the dialog box and scroll to the right. Then from the menu, select **Analysis→Forecast→Forecast Options** to open the dialog box again.

4. Configure the source data settings.

a) In the **Forecast Options** dialog box, under **Source Data**, in the **Ignore last** box, type ***0***

b) In the canvas, observe that the change in the forecast values includes the last month of source data.




## 5. Configure the forecast model settings.

a) In the **Forecast Options** dialog box, under **Forecast Model**, in the **Forecast Model** drop-down list, review the options. Tableau recommends using the **Automatic** setting for forecasts, unless you have a specific reason to change it.

b) Select **Custom**.

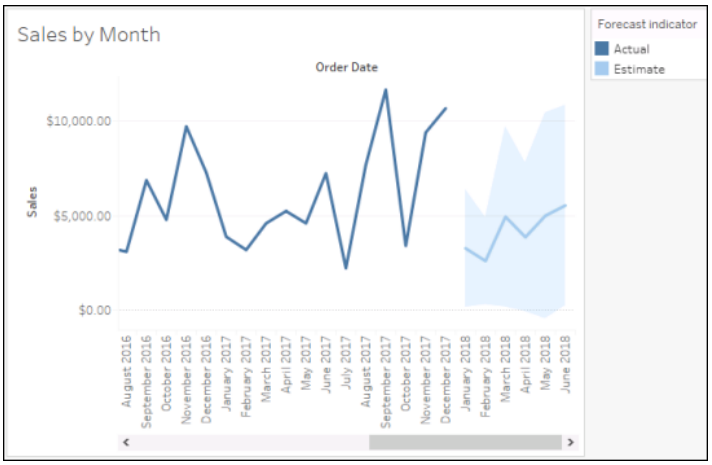
c) In the **Trend** and **Season** drop-down lists, try selecting different values and observe how the forecast changes.

d) In the **Forecast Model** drop-down list, select **Automatic**.

e) In the **Show prediction intervals** drop-down list, select **99%**.

f) Select **OK**.

g) In the canvas, observe how increasing the percentage of the prediction increases the range of the prediction interval.




## 6. Save the workbook.

a) From the menu, select **File→Save**.

b) From the menu, select **File→Close**.

# Summary

In this lesson, you plotted and analyzed the accuracy of trend lines and found residuals. You then created forecasts, adjusted forecast configurations, and reviewed forecast results.

## How familiar are you with statistical analysis? How will you use statistical analysis in Tableau?

## How will you use forecasting functions in Tableau?

**Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience.

# 8 Enriching Visualizations, Dashboards, and Maps

**Lesson Time: 2 hours, 30 minutes**

## Lesson Introduction

Tableau® has many features that allow you to customize and enrich visualizations, dashboards, and maps to help showcase the story told by the data. In this lesson, you will customize mapped data, and enhance visualizations and dashboards.

## Lesson Objectives

In this lesson, you will:

- Customize mapped data.
- Enhance visualizations with tooltips.
- Enhance dashboards with actions.

# TOPIC A

## Customize Mapped Data

Maps can create compelling visualizations, but you aren't limited to using pre-made maps or predefined areas, such as state or country boundaries. In this topic, you will learn how to customize maps to meet the needs of the story you wish to tell with your data.

### Unrecognized Locations

Tableau may be able to plot data on maps, but may not be able to uniquely identify data without providing additional information. For example, if you plot several cities across the United States, you will see the marks, but when you double-click to access them, much of the underlying information will be unknown because Tableau may be confused due to identical city names in other countries, such as Paris in France versus Paris in Texas.

These types of issues can often be solved by adding more fields, such as the next level in the geographic hierarchy, to **Details** on the **Marks** card. For example, if mapping cities in the United States, you might solve the aforementioned issue by dragging the State field to **Details** on the ***Marks card***. Tableau uses the additional data to create a hierarchy and resolve ambiguities.

Tableau will also place a notification of the number of unknowns in that lower-right corner of the map. You can click notification to hide it, filter the data, edit the locations, or display the data at a default position.

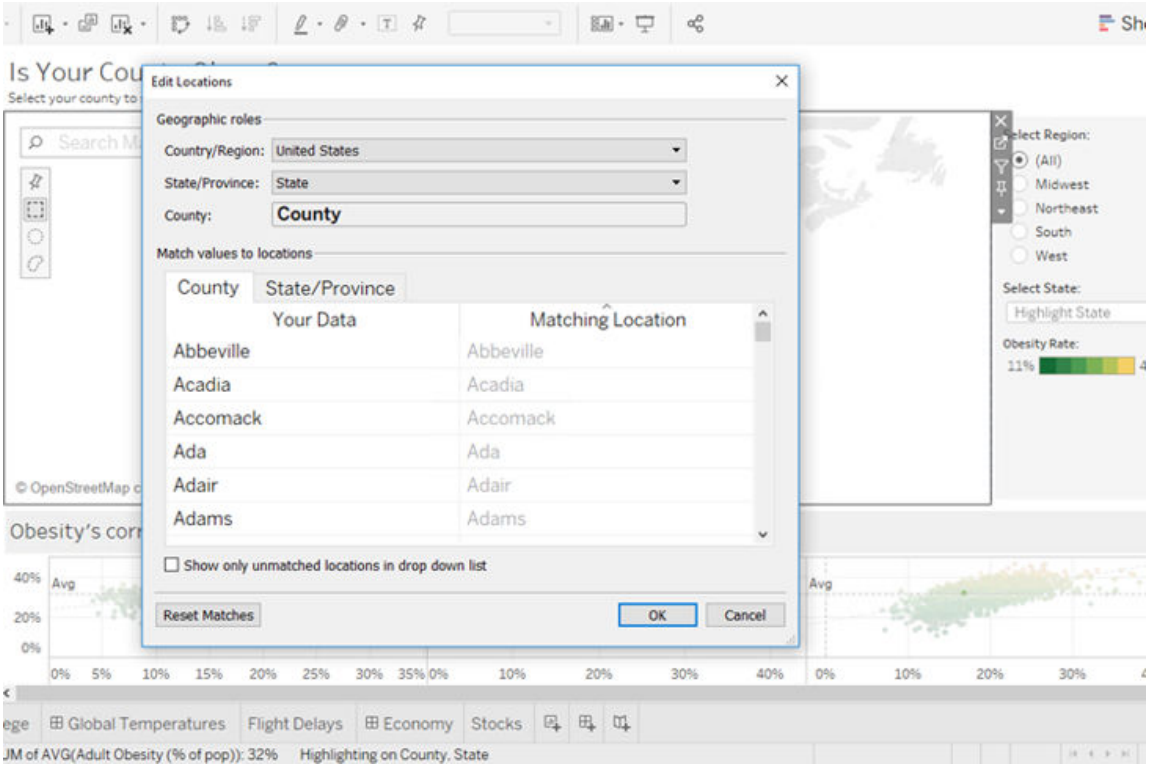



*Figure 8-1: Unrecognized locations.*

Unknowns might be the result of spelling mistakes in the data or territory updates. In those cases, you can choose to edit the unknowns. When you edit the unknowns, you can use the default for things like the country and region, or the state and province, or select a field Tableau will use to pull that data. You will also see any issues highlighted with a yellow triangle containing an exclamation

mark on the appropriate tab. You can select those tabs and correct the issues, such as selecting the correct entries for misspelled names, or entering the longitude and latitude for places Tableau cannot find.

## Batch Geocoding to Extended Mapping

If you have data sets that don't have geocoding information (such as longitude and latitude), you can add that data to use third-party geocoding software outside of Tableau, and then import the newly created geocoding data back into Tableau. For example, if you wanted to plot store locations, but only have the address of each store in the data, you can export the list of addresses using services,\ such as these:

- **csv2geo.com** for batch geocoding to get longitude and latitude from addresses.
- **Geocod.io** to for batch geocoding.
- U.S. Census geocoder available at **https://www.census.gov/geo/maps-data/data/geocoder.html**.

There are many free and fee-based services that provide batch geocoding services. Many services provide an address to geocode, while others provide information like ZIP Code or area code to geocode. You should review several services to find the best fit when looking for batch geocoding.

## Spatial Files

Spatial files are a file format developed by Autodesk® that contain geocoded data, such as longitude and latitude coordinates. These files were developed for use with Autodesk programs, such as MapGuide and AutoCAD Map 3D. Tableau can also connect to spatial files including Sharefiles, MapInfo tables, Keyhold Markup Language (KML) files, and GeoJSON files. To connect to spatial files, the folder with the file must contain all the supporting files for the type of file you're connecting to. You can connect to spatial data files as you would any other data source.

Some common places where you can find spatial files are:

- **Census.gov**
- **Data.gov**
- U.S. Energy Information Administration
- USGS Water Resources
- Geospatial Information Authority of Japan
- London Datastore
- Environment Geographic Information Systems South Africa

Once connected, you can join tables in spatial files and create maps from spatial files.

### Additional Information for Connecting to Spatial Files

For the following types of spatial files, you must have the accompanying files present in the folder of the data source you're connecting to:

- ESRI shapefiles must be accompanied by SHP, SHX, and DBF files.
- MapInfo tables must be accompanied by TAB, DAT, MAP, and/or ID, MID and MIF files.
- KML files only require the KML file.
- GeoJSON files only require the GEOJSON file.

## Geometry Field

The **Geometry** field is used to allow you to define boundaries on a map. Once you connect to spatial data, a **Geometry** field is created in Tableau for point geometries or polygons. Use the **Geometry** field to create your map using the spatial data. The values in the **Geometry** field show the geometric primitive or aliasing for the underlying spatial data. For example, **Points** for point

geometries, **Linestring** or **Multilinestring** for linear geometries, and **Polygon** or **Multipolygon** for polygons. But this is simply an alias for the underlying spatial data.

To build a map, double-click the **Geometry** field to add it to the **Marks** card. At the same time, the **Latitude (generated)** and **Longitude (generated)** fields are added to the **Columns** and **Rows** shelves, and a map view is created. You should consider filtering the number of geometries being added to the view in large spatial data sets to improve performance and speed up the map render time. You can then add levels of detail present in the spatial file, and customize the data as you would other maps.

## Custom Geocoding

There may be locations that Tableau can't map, such as street addresses. You can add these locations as custom geocoded locations. That essentially means you are assigning latitude and longitude coordinates to the locations so that Tableau can plot them. This also allows you to create custom ***geographic roles*** that you can use as you create map views in Tableau.

It's also possible to create maps without using spatial files or background images by importing custom geocoding information. You might have this available or may wish to create it so users can work with mapped data without having to work directly with longitude and latitude in the data. For example, if you created custom geocode data for each location of a business, you could import that and assign fields, such as Office and Country, to geocode roles. You can then plot the locations on the map by dragging those fields to **Details** on the **Marks** card.

To do this, you must do the following:

- Create one or more CSV files with your location data. These must contain **Latitude** and **Longitude** columns, and the values must be real numbers and include at least one decimal place.
- Optionally create a scheme.ini file to help Tableau properly interpret the data as geocoded data.
- Import the files into Tableau.
- Assign geographic roles to your data.

### Additional Information

For more information on geocoding data, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/maps_geographicroles.html**.

## Custom Geocoding Files

When adding custom geocoding to your data, you must create a CSV file for importing into Tableau. The contents of the CSV file are different based on whether you are extending an existing geographic hierarchy, adding a new one, or adding a new geographic role. Regardless, the CSV file must contain the hierarchy described as well as the **Latitude** and **Longitude** columns that contain real numbers and at least one decimal place.

For example, an import file for Country (Name), State/Province, and City would look like the following.

| *Country (Name)* | *State/Province* | *City* | *Latitude* | *Longitude* |
|---|---|---|---|---|
| United States | Oregon | Portland | 45.512231 | -122.658719 |
| United States | Oregon | Bend | 44.058173 | -121.315310 |

An import file containing information for Country (Name) and ZIP Code/Postal code would look like the following.

| *Country (Name)* | *ZIP Code/Postal Code* | *Latitude* | *Longitude* |
|---|---|---|---|
| United States | 97223 | 45.4423 | -122.7727 |

| Country (Name) | ZIP Code/Postal Code | Latitude | Longitude |
|---|---|---|---|
| United States | 97224 | 45.4077 | -122.8095 |

If you're importing roles, you need to include the new roles along with any parent roles, and longitude and latitude data. For example, to add Cascade Mountains in a role of Mountain, your import file might look like the following.

| Mountain | Country (Name) | State | Latitude | Longitude |
|---|---|---|---|---|
| St. Helens | United States | Washington | 46.1914 | -122.1956 |
| Hood | United States | Oregon | 45.3736 | -121.6960 |

### Additional Information

For additional information, see **https://onlinehelp.tableau.com/current/pro/desktop/en-us/help.htm#custom_geocoding.html**.

## Mapping with Background Images

You can add a static image, such as a Google Maps or Openstreet image, and plot your data over the image by selecting **Map→Background images**. The image you add should relate to the image being imported. For example, if you wanted to plot stops along bus lines, you import a map of the bus lines, and perhaps download a JPG image of the bus lines and stops, when you're connecting to the data set. Your data source can include **Line name** and **Stop name** information. You need to add X and Y coordinates for each stop. X and Y need to be measures, and a scale must be assigned related to the size of the background image.

When you choose to import a background image with the data set, you can see pixel locations for bitmaps. For example, if your image is 800 x 600 pixels, the X axis would run from a leftmost point of 0 to a rightmost point of 800, while the Y axis would run from a bottommost point of 0 to a topmost point of 600. That creates the scale along the lines of each access.

**Note:** If you're importing a map, you would enter longitude and latitude information rather than pixel information.

Once the scale is created, you can add X on the **Columns** shelf and Y on the **Rows** shelf. You can then pick any point of interest on the image, such as a stop, and select **Annotate→Point**. That will give you the coordinates on the image or map. You can drag the marker to the next point of interest to quickly capture coordinates for the new location. You can then add that information to your data source. Once you've entered all the information, refresh the data source, and the stop and line information will be plotted on the image. You can then customize the image color, size, and more to make plotted points work well with the imported images.

## Options for Joining or Blending Geocode Data

You can join or blend geocode data with other data as long as you have related fields you can map in each data source. For example, if you have sales data for each store in your organizations, and that sales data includes the store name and address, you can create a geocoded file that contains the coordinates of the address in question to precisely plot the store locations on a map. You could also join the data to U.S. Census data. This will let you find data about family size and median income in the areas around your stores, and find any insights related to sales and profitability that might be related to those metrics.

**Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Analyze Data with Advanced Mapping Features.**

# ACTIVITY 8-1

## Analyzing Data with Advanced Mapping Features

### Data File

C:\095205Data\Enriching Visualizations, Dashboards, and Maps\MyFootPrintSports_2018_8.twb

### Before You Begin

Tableau Desktop is open.

### Scenario

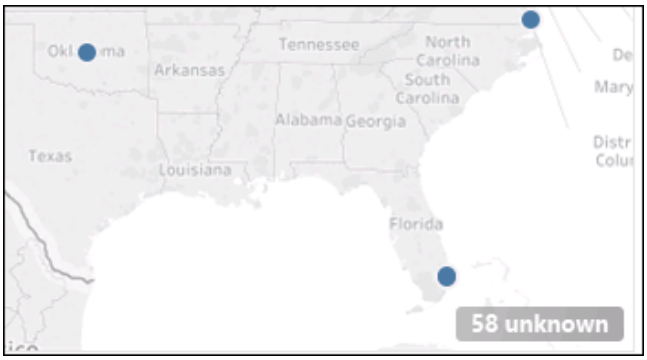
When another user tried creating their own map in Tableau by adding the city, they received a message about there being unknown locations. You decide to investigate the issue and find the source of the problem. After that, you want a map that shows sales by category using color and size. You also want to see profit by size, using a gray color on the same map.

---

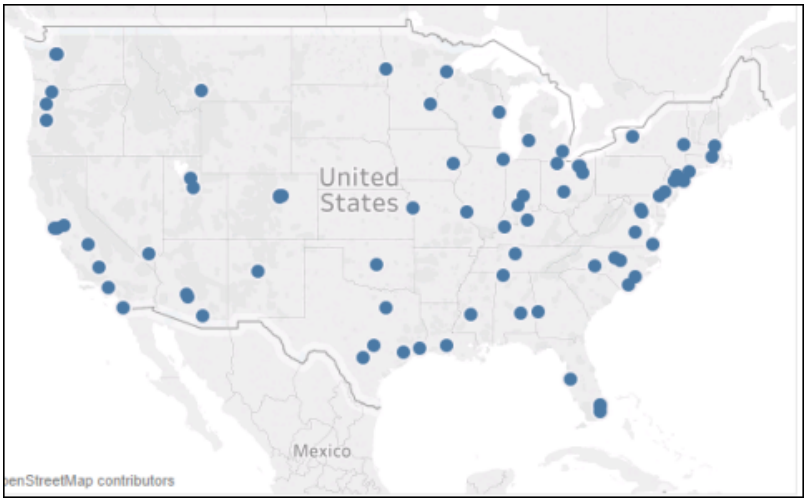
1. Open the worksheet for the lesson.
   a) From the menu, select **File→Open**.
   b) In the **Open** dialog box, browse to the **C:\095205Data\Enriching Visualizations, Dashboards, and Maps** folder.
   c) Select **MyFootPrintSports_2018_8** and select **Open**.

2. Save the workbook.
   a) From the menu, select **File→Save As**.
   b) In the **Save As** dialog box, verify you are in the **C:\095205Data\Enriching Visualizations, Dashboards, and Maps** folder.
   c) In the **File name** box, type *MyCopy_MyFootPrintSports_2018_8* and select **Save**.

3. Review unknown cities.
   a) On the toolbar, select the **New Worksheet** button.
   b) On the **Sheet 14 worksheet** tab, in the **Data** pane, under **Dimensions→Location**, select and drag **City** to the canvas.
   c) Observe that Tableau identified 58 cities that it could not resolve.

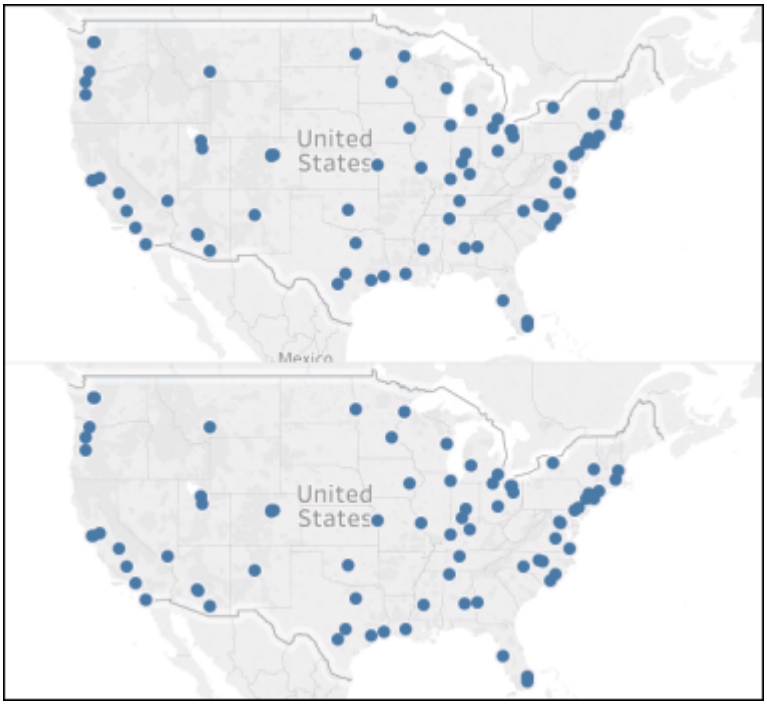


d) Select **58 Unknown**.
e) In the **Special Values for [City]** dialog box, select **Edit Locations**.
f) In the **Edit Locations** dialog box, review the list of unknown locations.
g) Observe that the **State/Province** drop-down list is set to **None**.
h) Select **Cancel**.
i) In the **Data** pane, under **Dimensions→Location**, select and drag **State** to the canvas.
j) Observe that the unknown cities have been resolved. This is because Tableau needed the state to put those locations into context. It is most likely that those city names exist in multiple countries and/or states.




## 4. Create a dual axis map.

a) In the **Data** pane, under **Measures**, select and drag **Latitude (generated)** to the **Rows** box to the right of the existing **Latitude (generated)**.
b) Observe that there are two maps in your view.

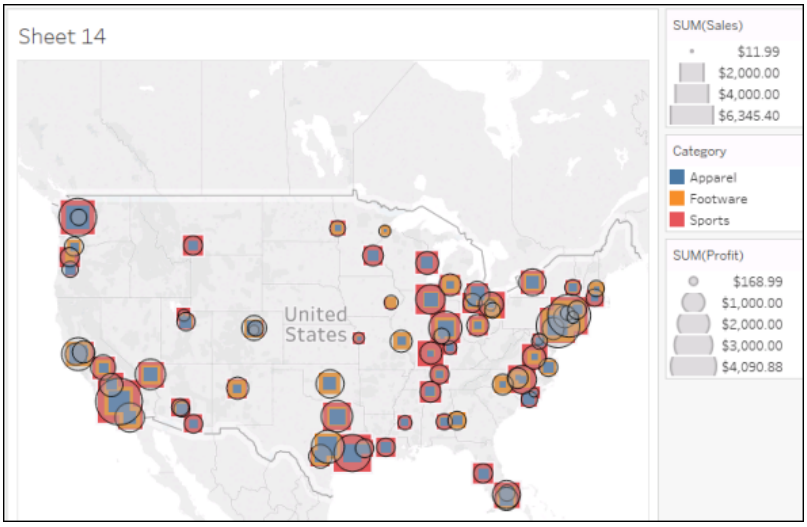



c) In the **Marks** card, select the first **Latitude (generated)** section.

d) In the **Data** pane, under **Dimensions→Products**, select and drag **Category** to the **Color** box in the **Marks** card.
e) In the **Data** pane, under **Measures→Orders**, select and drag **Sales** to the **Size** box in the **Marks** card.
f) In the **Marks** card, in the **Automatic** drop-down list, select **Square**.
g) In the **Marks** card, select **Size**.
h) Move the slider to the halfway point.

5. Format the second axis of the map.

a) In the **Marks** card, select the second **Latitude (generated)** section.
b) In the **Data** pane, under **Measures→Orders**, select and drag **Profit** to the **Size** box in the **Marks** card.
c) In the **Marks** card, select **Size**.
d) Move the slider to the halfway point.
e) In the **Marks** card, select **Color**.
f) In the **Color menu**, select a light gray color.
g) Change **Opacity** to **30%**.
h) in the **Border** drop-down list, select black or another dark color.
i) In the **Rows** box, right-click the second **Latitude (generated)** and select **Dual Axis**.
j) Observe the two overlaid maps with one showing sales by region and the other showing sales by category.




6. Rename the worksheet.

a) Right-click the **Sheet 14** tab and select **Rename**.
b) Type ***Dual Axis Sales Map*** and press **Enter**.

# ACTIVITY 8-2
## Mapping with Custom Backgrounds

### Before You Begin
The **MyCopy_MyFootPrintSports_2018_8** workbook is open in Tableau Desktop.

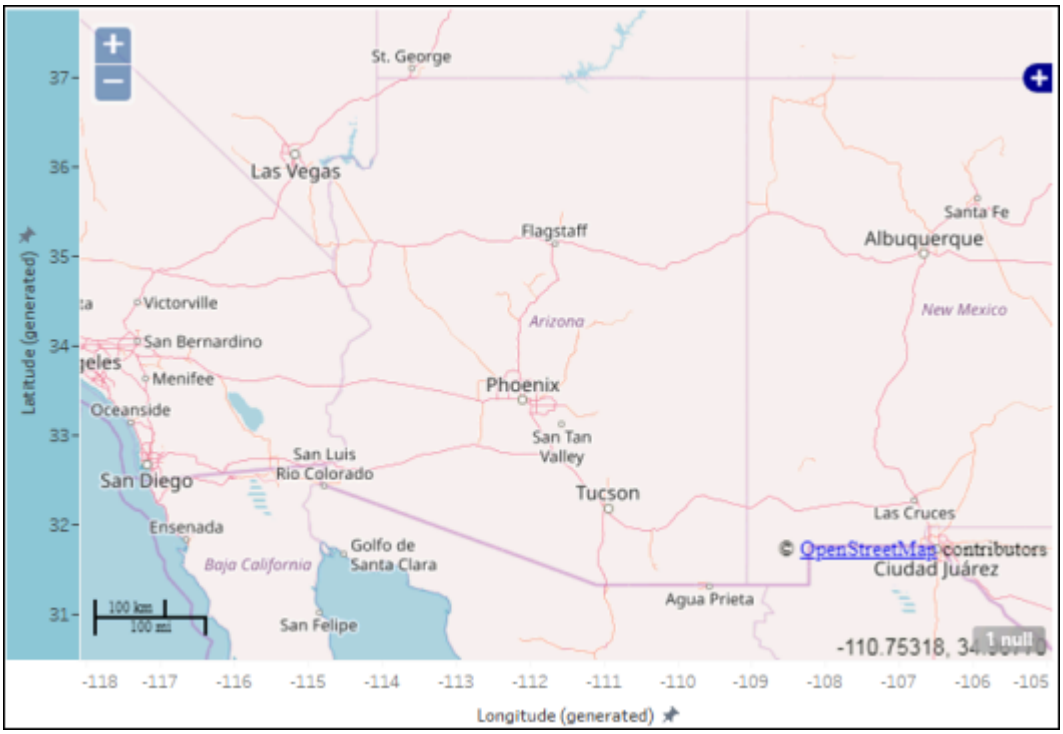
### Scenario
You want to use a custom map for some of the areas you have sales for. You decide to use Arizona as a test to configure a custom background. You have already obtained the image and the latitude and longitude coordinates. Now you will add them to Tableau.

---

1. Add the background image to Tableau.
   a) On the toolbar, select the **New Worksheet** button.
   b) On the **Sheet 15** worksheet tab, from the menu, select **Map→Background Images→Orders+ (MyFootPrintSportsSalesProducts_2018)**.
   c) In the **Background Images [Sheet 15]** dialog box, select **Add Image**.
   d) In the **Add Background Image** dialog box, select **Browse**.
   e) Navigate to **C:\095205Data\Enriching Visualizations, Dashboards, and Maps**, select **Arizona**, and select **Open**.
   f) In the **X Field** drop-down list, select **Longitude (generated)**.
   g) In the **Left** box, type *-118.09204*
   h) In the **Right** box, type *-104.93042*
   i) In the **Y Field** drop-down list, select **Latitude (generated)**.
   j) In the **Bottom** box, type *30.47782*
   k) In the **Top** box, type *37.74533*
   l) Select **OK** twice.

2. Configure the latitude and longitude for the background image.
   a) In the menu, select **Map→Background Maps→None**.
   b) In the **Data** pane, under **Measures**, double-click **Latitude generated**.
   c) In the **Data** pane, under **Measures**, double-click **Longitude generated**.
   d) In the canvas, right-click the **Longitude** axis and select **Edit Axis**.
   e) In the **Edit Axis [Longitude generated]** dialog box, under **Range**, select **Fixed**.
   f) In the **Fixed start** box, type *-118.09204*
   g) In the **Fixed end** box, type *-104.93042*
   h) Close the **Edit Axis [Longitude generated]** dialog box.
   i) Right-click the **Latitude** axis and select **Edit Axis**.
   j) In the **Edit Axis [Latitude generated]** dialog box, under **Range**, select **Fixed**.
   k) in the **Fixed start** box, type *30.47782*
   l) In the **Fixed end** box, type *37.74533*
   m) Close the **Edit Axis [Latitude generated]** dialog box.

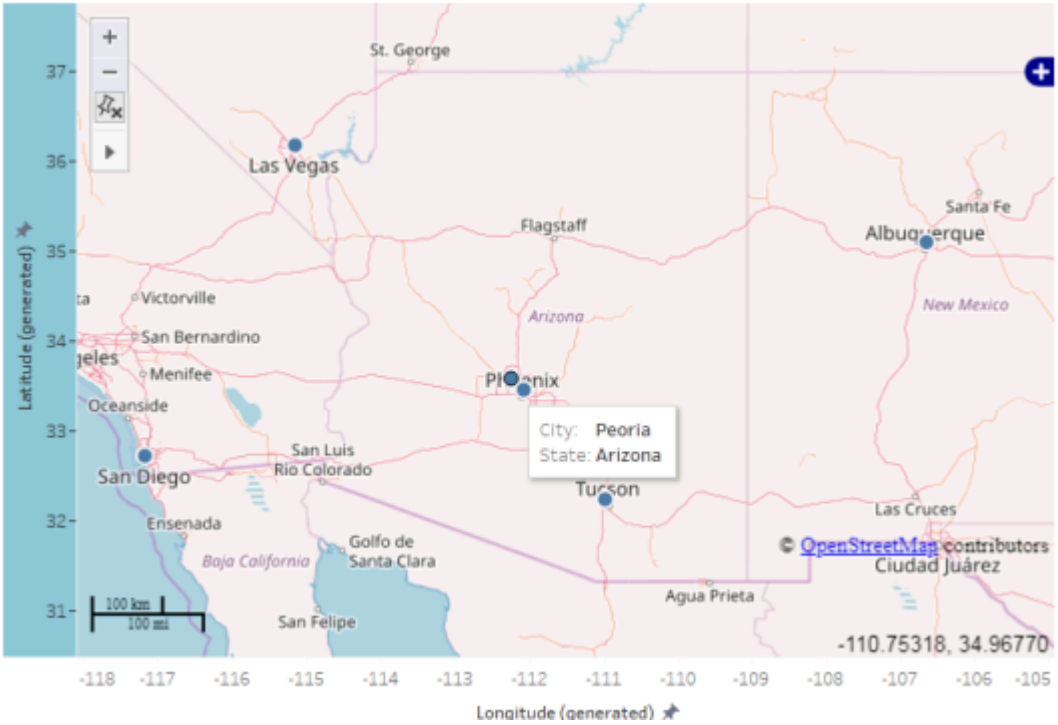
n) Observe the Arizona map in the view.




3. Use the map.

a) In the **Data** pane, under **Dimensions→Location**, double-click **City**.

b) Observe the Arizona map with city marks from your data.




4. Rename the worksheet.

a) Right-click the **Sheet 15** tab and select **Rename**.
b) Type ***Arizona Map*** and press **Enter**.

# TOPIC B

## Enhance Visualizations with Tooltips

Tooltips provide basic information about the visualization on the sheet, but they can be customized to show much more. In this topic, you will use tooltips to enrich visualizations, help tell the story showcased by the data, and provide additional information.

### Stories in Tableau

In Tableau, a ***story*** is a sequence of sheets or dashboards that work together to step your audience through a series of views in sequence to showcase a specific set of insights or how you came to a specific conclusion. You can create stories to tell a data-driven narrative, provide context, demonstrate how decisions relate to outcomes, or to simply make a compelling case. Some examples of stories are:

- Relating how driving laws and technologies helped reduce automobile fatalities from 1950 to the present day.
- Charting out technological breakthroughs that helped mankind go to the moon.
- Describing how the analysis you're working on helped you arrive at the conclusion you found.

Similar to dashboards, when you create a story, the **Data** pane is replaced by the ***Story pane***, which shows you the sheets and dashboards available in the workbook. You can also drag text blocks to the story, select to show a title, and set the size of the story.

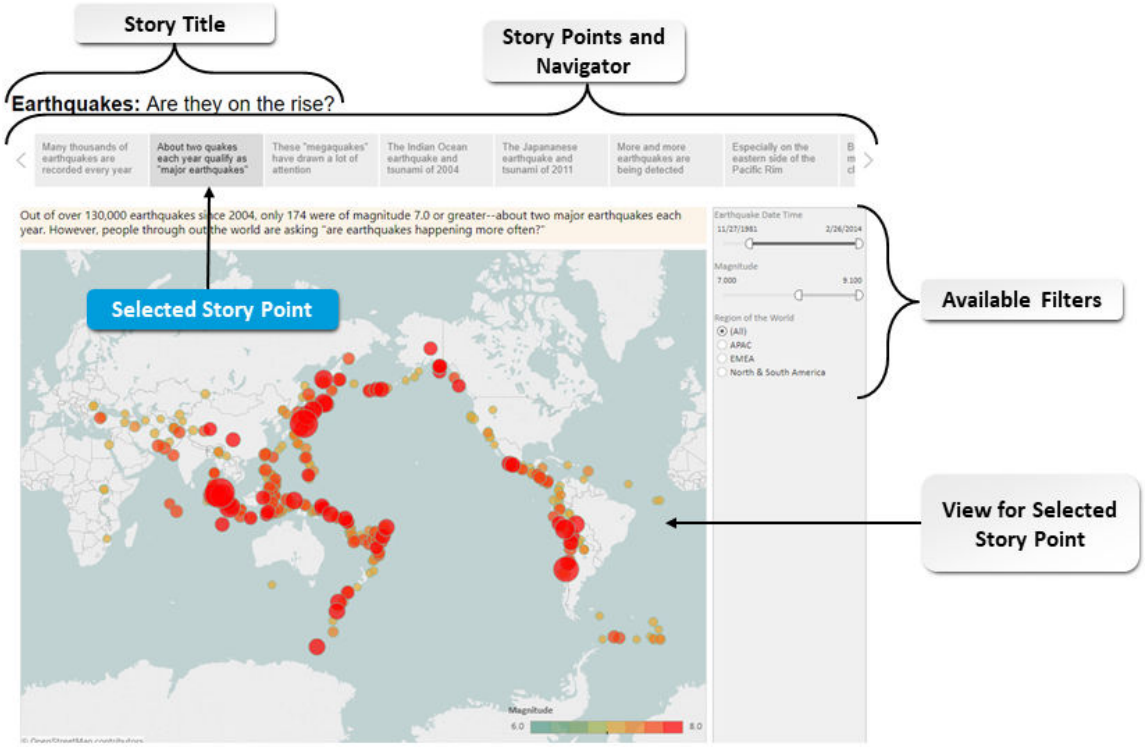



*Figure 8-2: Stories in Tableau.*

Stories can be shared to Tableau Public, Tableau Server, or Tableau Online. Once they are, users can interact with the story and explore the data behind it.

**Note:** To learn more about creating stories, check out the Spotlight on **Best Practices for Creating Stories** presentation from the **Spotlight** tile on the CHOICE Course screen.

## Storytelling with Tooltips

By default *tooltips* provide some information and context about marks in the view, such as the underlying fields used to make the view. However, tooltips can be heavily customized. You can add text to provide context for what the marks in the view show. You can format and color that text to emphasize specific elements in the view. You can add information from fields so the context provided relates specifically to the mark being examined. You can even add fields to tooltips that aren't in the view, or show a different visualization from another worksheet in a tooltip.

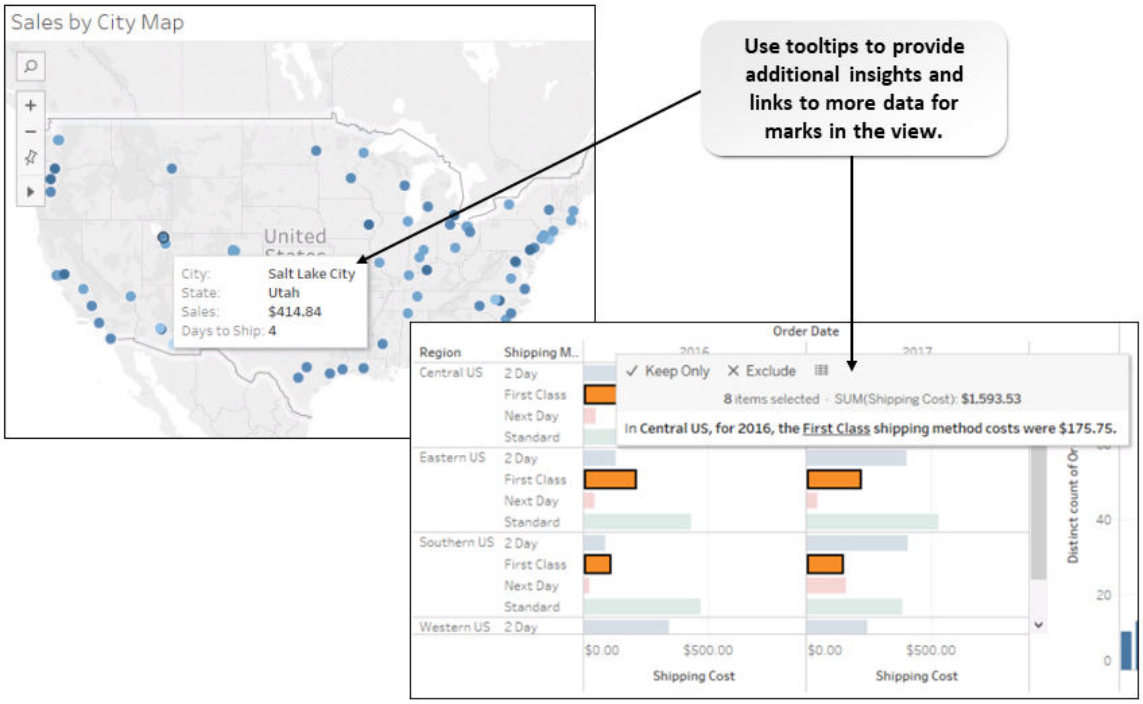



*Figure 8-3: Storytelling with tooltips.*

When creating visualizations, you should consider how tooltips can help you tell your story. For example, you might have a bar chart showing product profitability. If users frequently ask about the sales figures that drove that profitability, you can include that in a tooltip. In another example, you may prefer to show surface data in your dashboard, but find that adding a sheet with the detail makes the dashboard too busy. To solve the problem, you could move that detailed data to a secondary visualization accessed through a tooltip.

The power and flexibility of tooltips can enhance visualizations, and provides an extended range of options to help you tell your story and make additional information available to users.

## Edit Tooltips

Tooltips are configured at the worksheet level. To customize them, select **Edit→Tooltip**.

The **Edit Tooltip** dialog box has the following options.

| *Configuration Option* | *Description* |
|---|---|
| **Text formatting** | These are standard text formatting controls that can be used to enhance text in the text area by setting font, size, color, bolding, and so forth. |

| Configuration Option | Description |
|---|---|
| **Insert button** | This allows you to insert any field from the view, or any field that has been dragged from the tooltip shelf to the text area. Fields added must have the field name surrounded by brackets, such as <Field Name>. The text displayed is in the context of the selected mark and can be formatted like other text in the text area. |
| **Text area** | Type any text and add fields from the view or from the tooltip shelf. You can format the text and field output with the tools in the text formatting area. |
| **Show tooltips** | Uncheck this box to stop showing tooltips on the sheet. Leave the box checked and select from the drop-down list to set how quickly tooltips will appear. |
| **Include command buttons** | The command buttons appear at the top of the tooltip and allow users to include or exclude marks, make groups and sets, and view the underlying data for the marks. You can turn off the command buttons by essentially removing the command bar when you uncheck this box. |
| **Allow selection by category** | If this box is checked, you can select marks in the view that have the same value. For this to work, a field must have been inserted into the text area. If this box is checked, fields inserted in this way will be underlined, similar to a hyperlink in the tooltip. You can click them to highlight like data. For example, if you are looking at a bar chart showing combined quarterly sales and the tooltip shows the current quarter, such as Q2, you could click the Q2 object to highlight all Q2 data in the view. |

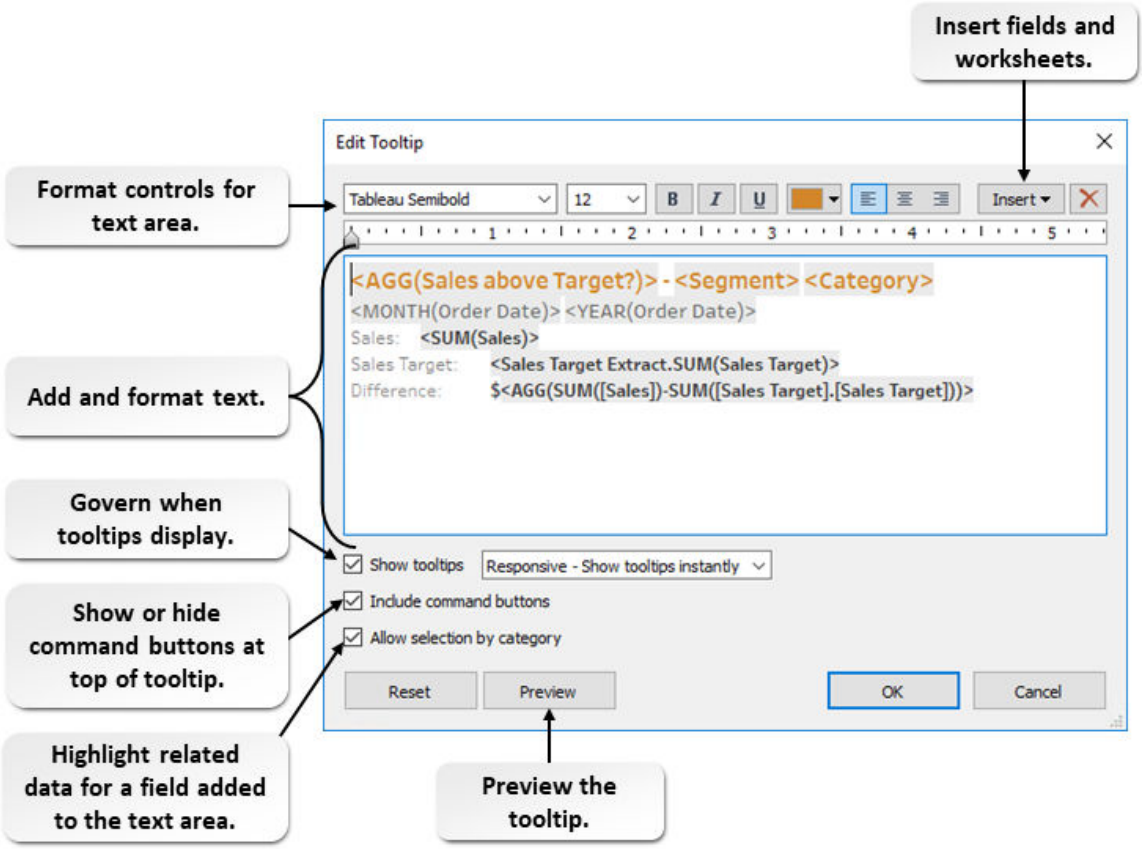



*Figure 8–4: The Edit Tooltip dialog box.*

## Conditional Tooltips

You can create conditional tooltips using calculated fields. For example, when doing an analysis of product sales, if you want the tooltip to show a clear message stating that a particular product is a new product, an existing product, or a discontinued product, you can create calculated fields to do that. Create a simple calculated field for each option using a logical IF-THEN-ELSE statement, such as the following:

```
IF [Product Cycle] = "New"
THEN [Product Cycle]
ELSE ""
END
```

This calculation returns true if the [Product Cycle] is "New." If it is, it returns the product cycle text of "New." If the [Product Cycle] is not "New," it returns nothing as the ELSE statement has nothing between the two quotation marks. You can name this calculated field [New].

You would create similar calculated fields for the [Product Cycle] of "Existing," and "Discontinued," and name the fields accordingly.

In the tooltip, you can include statements from each of the three calculated fields to have the tooltip display a message about the product cycle, such as:

```
This product is <ATTR(New)><ATTR(Existing)><ATTR(Discontinued)>.
```

Since only one of the product cycles can be true for each product, only one statement will return a result. The other statements will be false and return nothing. The ATTR ensures the row is unique. You can also use text formatting to color-code each product cycle differently, if you so choose.

## Viz in Tooltips

The Viz in Tooltips feature allows you to include an entirely separate visualization inside of a tooltip. This secondary visualization can show data related to the current or primary source visualization that helps explain or expand on the primary data. For example, if the primary source visualization shows sales for each category of products, the secondary visualization might show a pie chart that shows the top sellers in that category of products. Using the two together, you can see which categories produce the most sales, and then hover over any category you're interested in to see the tooltip with the secondary visualization that shows best sellers within that category.

To create a Viz in a tooltip, you have to create separate sheets for the primary source visualization and any secondary visualizations. Once created, return to the primary source visualization and add the Viz into a tooltip by selecting **Insert→Sheets**, and then selecting the target sheet with the visualization that you wish to include. The tooltip will display the visualization from the selected sheet. Viz in Tooltips visualizations are static in the Viz, and cannot be interacted with.

When a Viz in Tooltips is created to show a secondary visualization from a target sheet, a filter is created on the target sheet that filters that sheets view. Consequentially, the data is displayed in the tooltip based on the mark that is being hovered over in the primary source visualization. The filter is based on all fields by default. For filtering to work, the primary source and target sheets must share the same single data source. Also, a target sheet can only be used in a tooltip for one source sheet at a time. However, multiple target sheets can appear in the same tooltip. Only sheets can be included in tooltips, not dashboards or stories.

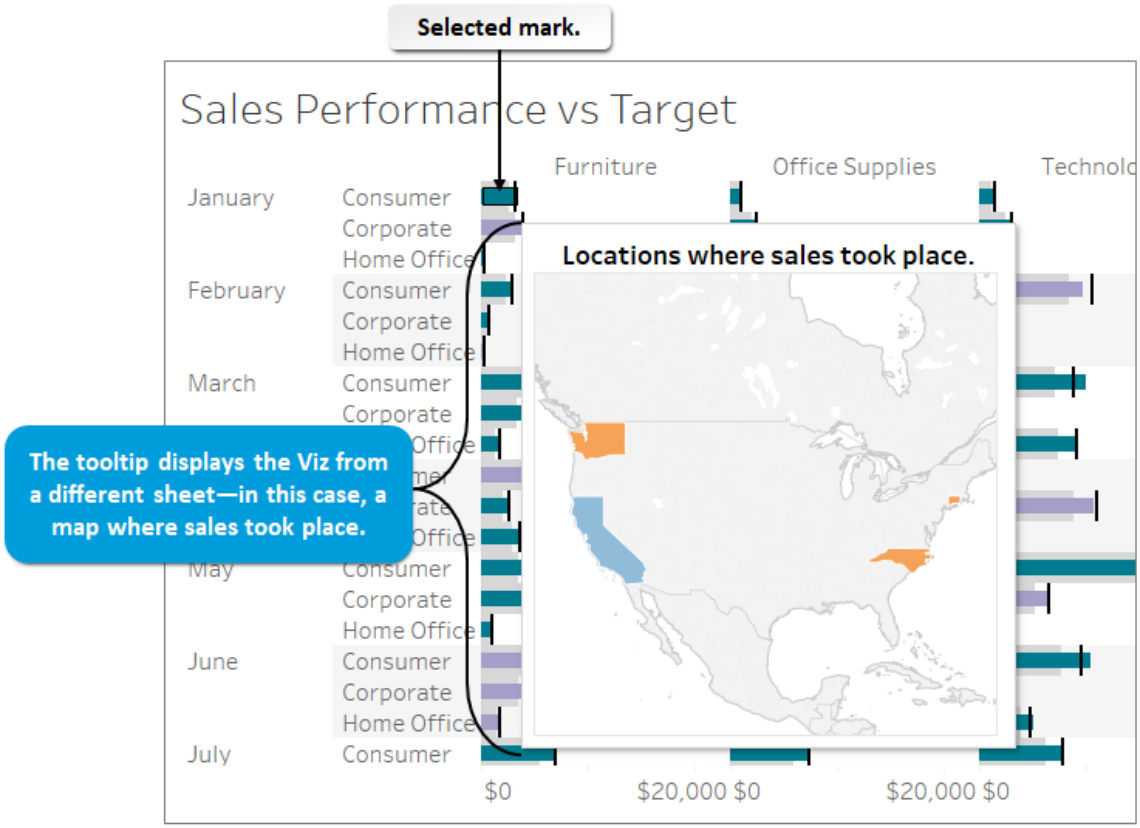



*Figure 8-5: The Viz in Tooltips.*

You can customize the tooltip display with Viz in Tooltips the same way as other tooltips. You can insert text and field names, turn off command buttons, and so forth. Some modifications to the secondary visualization, such as turning off axis labels, must take place on the target sheet where the secondary visualization is coming from.

## Viz in Tooltips Size

You may run into issues getting secondary visualizations to display correctly or show all of the data you desire when displayed in the tooltip. Some secondary visualizations may also be too large to be displayed; in this case, when you open the tooltip, you will receive the message "View is too large to show."

When you add a target sheet to a tooltip, the maxwidth and maxheight values are set at 300 pixels each. You can adjust these values by typing new ones in an embedded sheet statement. Tableau recommends not going over 600 pixels as that may obscure the primary source visualization. You can also adjust the view on the source data sheet by selecting **Entire View**, showing more data from the target visualization in the tooltip.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Enhance Visualizations with Tooltips.** |

# ACTIVITY 8-3

## Enhancing Visualizations with Tooltips

### Before You Begin

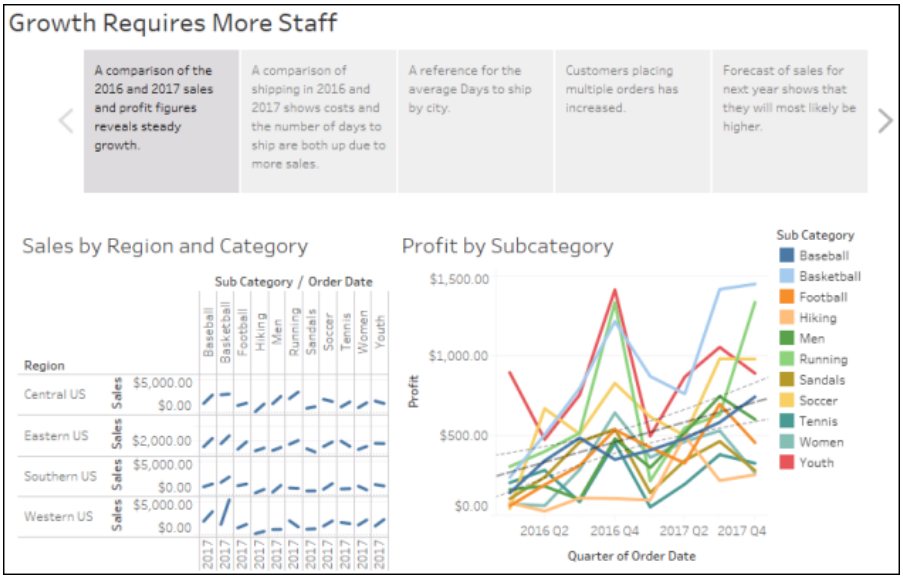
The **MyCopy_MyFootPrintSports_2018_8** workbook is open in Tableau Desktop.

### Scenario

You believe you've found solid data showing there has been substantial sales growth in the last two years, and you have created a story to help illustrate that the growth requires more internal resources to handle the increased load. Before you share your story, you want to enhance the worksheets used in the story with some changes to the tooltips to provide better information.

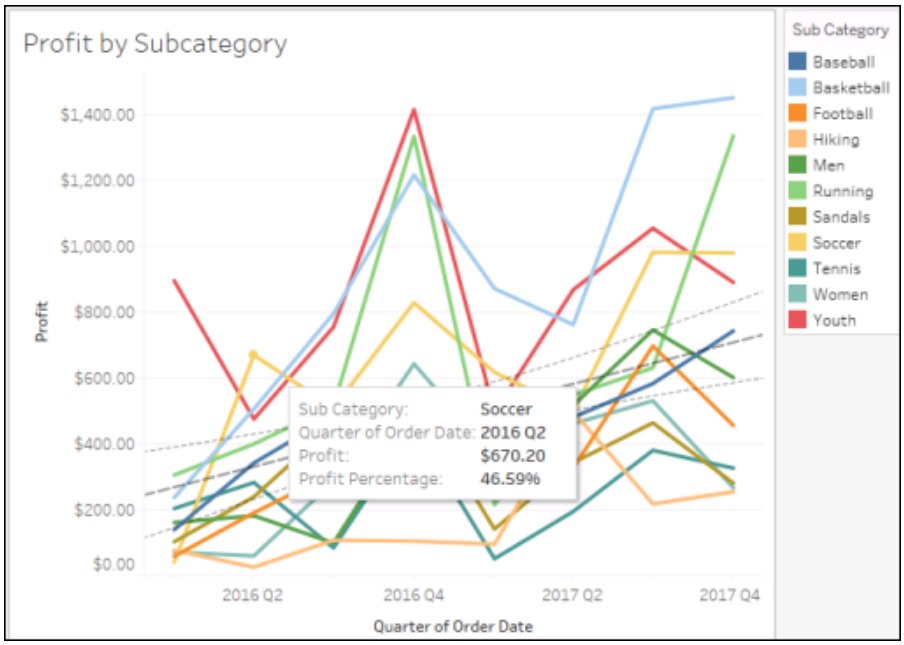
**Note:** Stories were covered previously in the *Tableau® Desktop: Part 1* course.

1. Review the story.
   a) At the bottom of the screen, select the **Growth Requires More Staff** story.
   b) Select and review each of the story points in the story to become familiar with the idea that the company has better sales, but is growing and needs more staff to meet demands.




2. Review the existing **Profit by Subcategory** tooltip.
   a) At the bottom of the screen, select the **Profit by Subcategory** worksheet.

b) In the canvas, hover the mouse over one of the lines of the chart and observe the tooltip created by default.

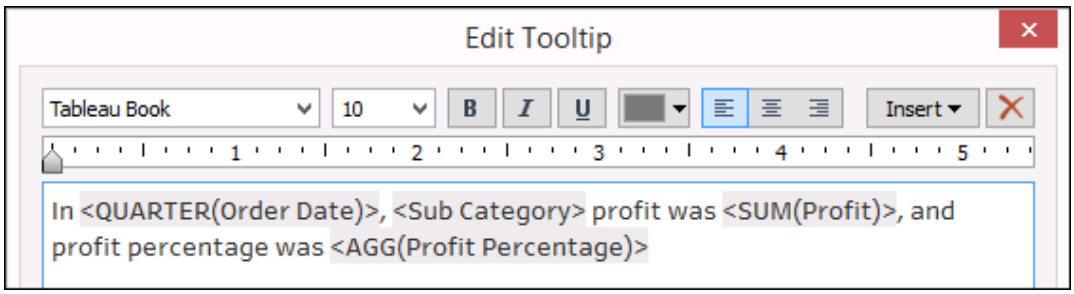



3. Edit a tooltip for a worksheet in the story.

a) From the menu, select **Worksheet→Tooltip**.

b) In the **Edit Tooltip** dialog box, in the text box, highlight and delete all of the existing text.

c) Re-create the tooltip text in the following image. You can type all of the text manually, or you can type the plain text manually and use the **Insert** drop-down menu to insert the four different fields.

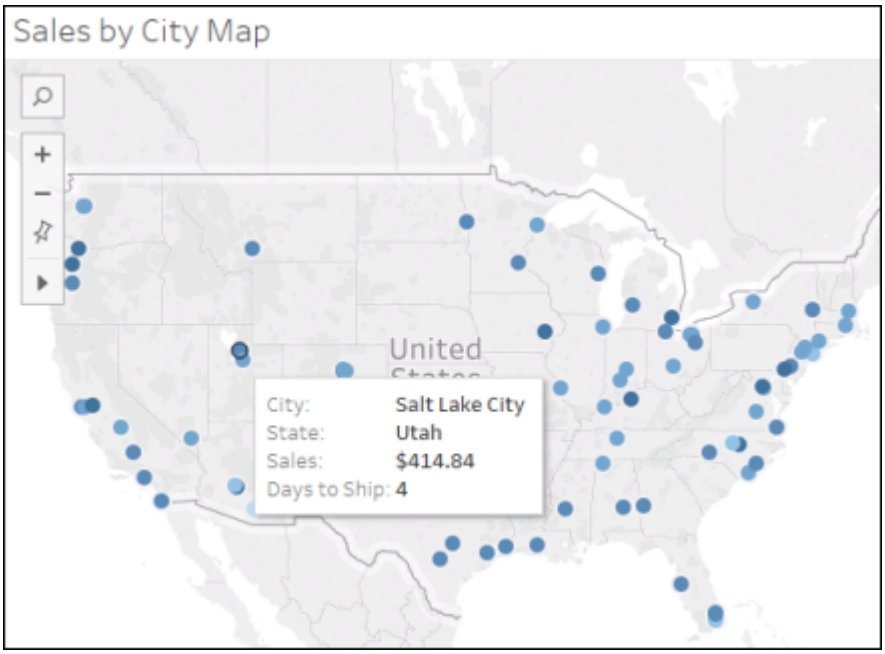



d) Select **OK**.

4. Review the existing **Sales by City Map** tooltip.

a) At the bottom of the screen, select the **Sales by City Map** worksheet.

b) In the canvas, hover the mouse over a city mark and observe the tooltip created by default.



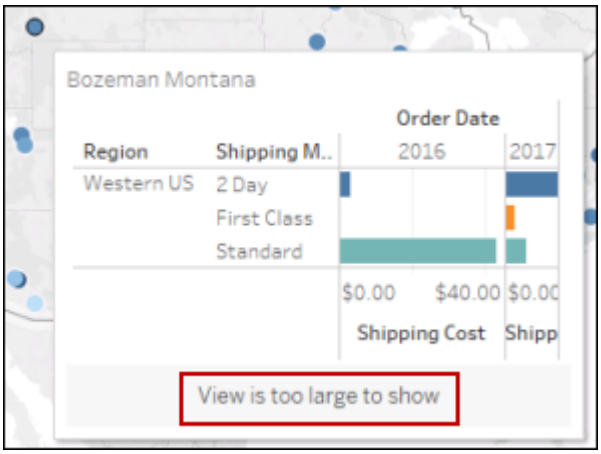

## 5. Create a Viz in a tooltip.

a) At the bottom of the screen, select the **Sales by City Map** worksheet.
b) From the menu, select **Worksheet→Tooltip**.
c) In the **Edit Tooltip** dialog box, in the text box, delete all of the existing text.
d) In the **Insert** drop-down list, select **City**.
e) In the text box, add a space after **<City>**.
f) In the **Insert** drop-down list, select **State**.
g) In the text box, press **Enter** to create a new blank line.
h) In the **Insert** drop-down list, select **Sheets→Shipping Cost by Shipping Method**.

```
<City> <State>
<Sheet name="Shipping Cost by Shipping Method" maxwidth="300"
maxheight="300" filter="<All Fields>">
```

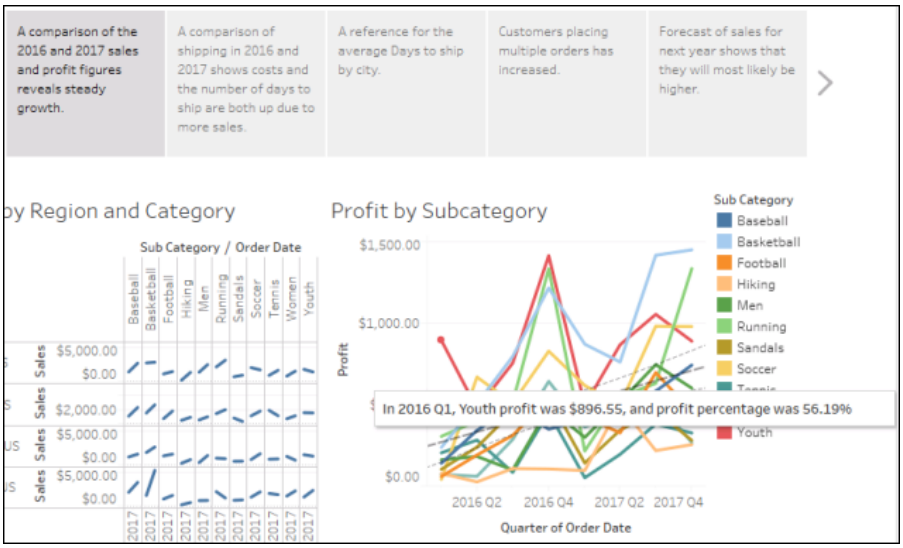
i) Select **OK**.

j) In the canvas, hover over a city mark and observe the tooltip and the message that the view is too large to show.



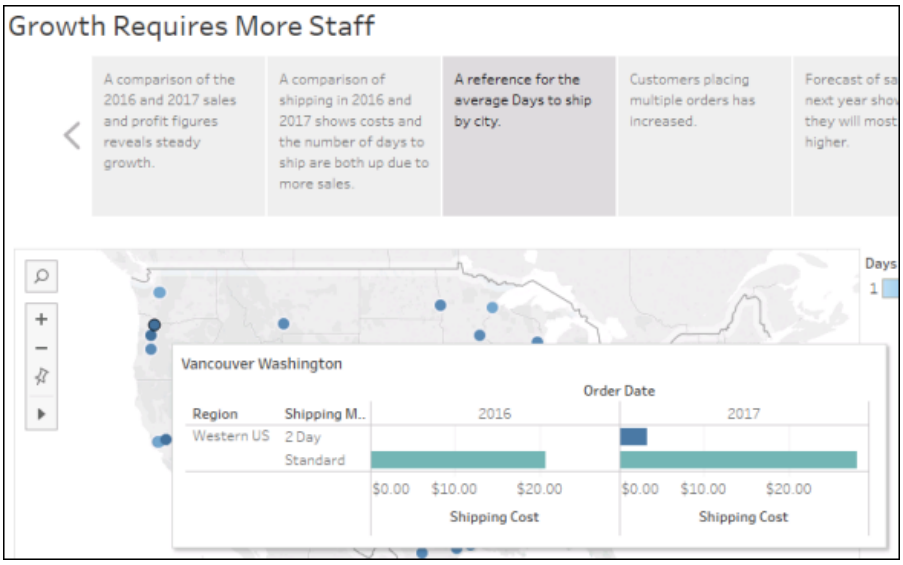

6. Adjust the updated tooltip.
   a) From the menu, select **Worksheet→Tooltip**.
   b) In the **Edit Tooltip** dialog box, in the text box, change **maxwidth="300"** to **maxwidth="600"**.
   c) Select **OK**.

7. Review the **Profit by Subcategory** worksheet in the story.
   a) At the bottom of the screen, select the **Growth Requires More Staff** story.
   b) On the first story point page, under **Profit by Subcategory**, hover the mouse over the lines to see the tooltip.



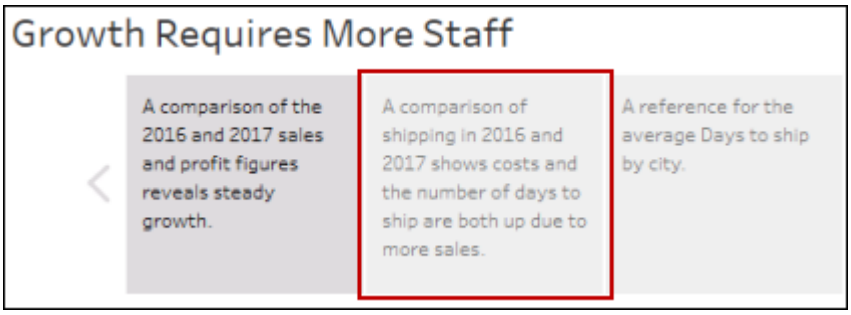

8. Review the **Sales by City Map** worksheet in the story.
   a) At the top of the screen, under **Growth Requires More Staff**, select the third story point.

b) On the first story point page, under **Profit by Subcategory**, hover the mouse over the lines to see the tooltip.



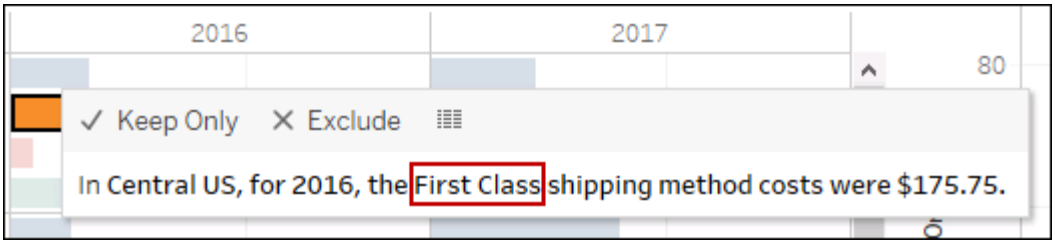

9. Review the discrete fields in tooltips in the **Shipping Cost by Shipping Method** worksheet in the story to highlight marks.

a) At the top of the screen, under **Growth Requires More Staff**, select the second story point.



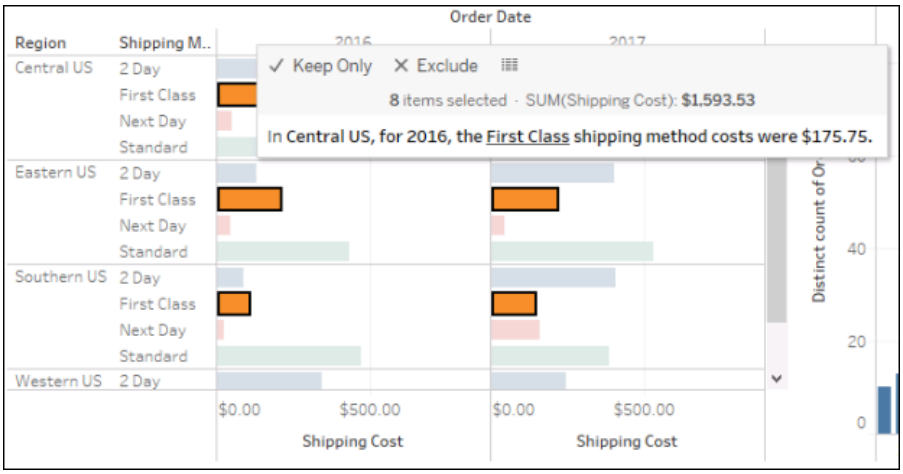
b) Under **Shipping Cost by Shipping Method**, in the **2016** column, for the **Central US** region, select the **First Class** mark and keep the mouse over the mark.

c) In the tooltip, select **First Class** and keep the mouse over the tooltip.



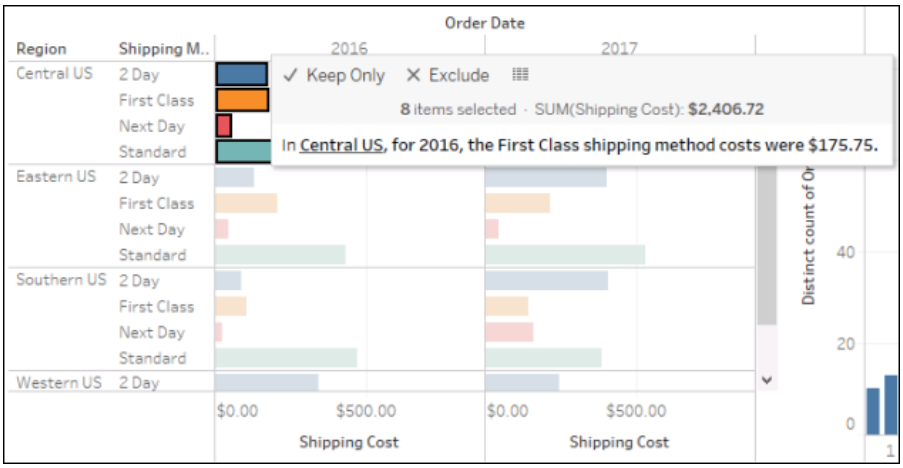

| | |
|---|---|
| | **Note:** Select the **First Class** text in the tooltip. If you keep your mouse over the mark or the tooltip after you select the mark, the tooltip will stay in place, allowing you to select the text inside the tooltip. |

d) In the canvas, observe that selecting the First Class shipping method in the tooltip highlighted all the marks for that shipping method.




e) In the tooltip, select **Central US** and keep the mouse over the tooltip.

f) In the canvas, observe that selecting the Central U.S. region in the tooltip highlighted all the marks for that region.




g) In the tooltip, select **2016** and keep the mouse over the tooltip.

h) In the canvas, observe that selecting the year 2016 in the tooltip highlighted all the marks for that year.

i) In the canvas, near the selected marks, select the white space to deselect the marks.

# TOPIC C

## Enhance Dashboards with Actions

Dashboards use data across a variety of visualizations to paint a picture that leads to insightful analysis. In this topic, you will use dashboard actions to filter and highlight information across dashboard elements, and expand the information available by providing URLs to external data.

### Dashboard Actions

***Dashboard actions*** allow you to add interactive elements to dashboards that show relationships in data between dashboard components, change the view to accentuate specific components, or to provide additional information. Actions are driven from the worksheet and can include dynamic values from fields in the view. Actions are activated in the view by a configured mouse action, such as a hover, selection (click), or from the menu of a tooltip.

Create an action by selecting **Dashboard→Action.** You can configure the following three types of actions:

- Highlight actions to highlight data in the view based on marks.
- Filter actions to filter data in the view based on marks.
- URL actions to provide additional information about marks in the view.

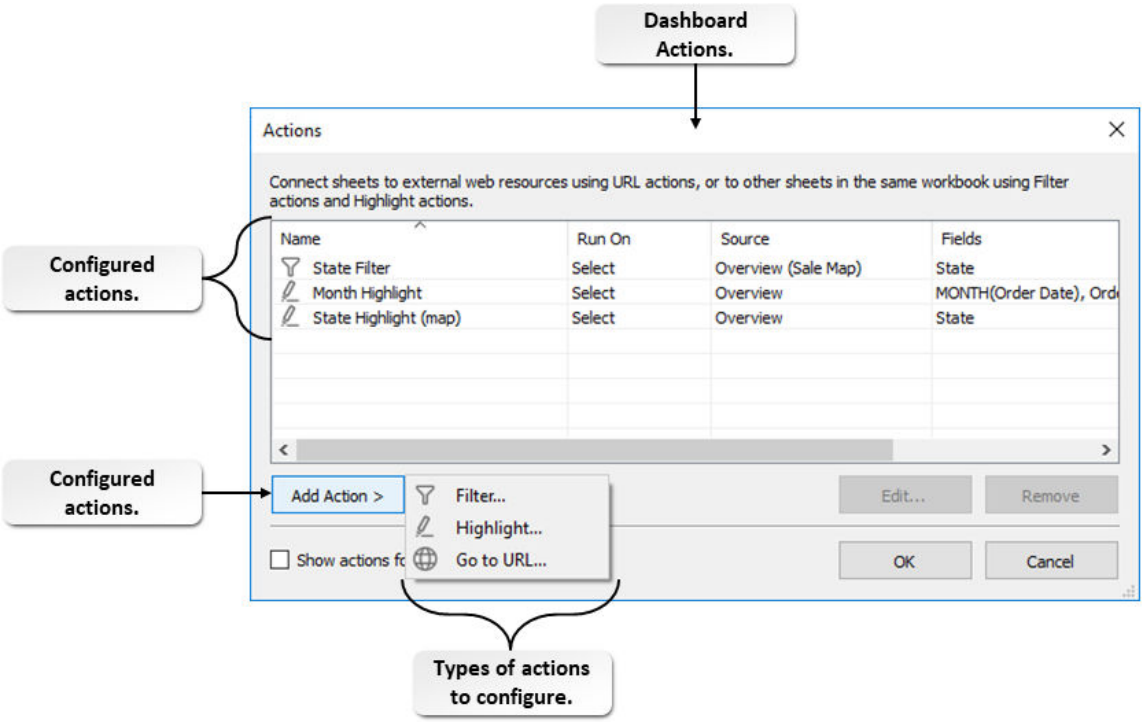



*Figure 8-6: Dashboard actions.*

| | |
|---|---|
| | **Note:** To learn more about creating dashboards, check out the Spotlight on **Best Practices for Creating Dashboards** presentation from the **Spotlight** tile on the CHOICE Course screen. |

## Name and Run Action

When configuring an action, you must always give the action a name, and specify the run action that tells Tableau when to activate the action. The choices are:

- **Select**: This is the default. The action is activated by clicking the mark.
- **Hover**: The action is activated by hovering over the mark.
- **Menu**: The action is included in the tooltip menu as an action in tooltip and can be clicked inside the tooltip. This is common for URL actions.

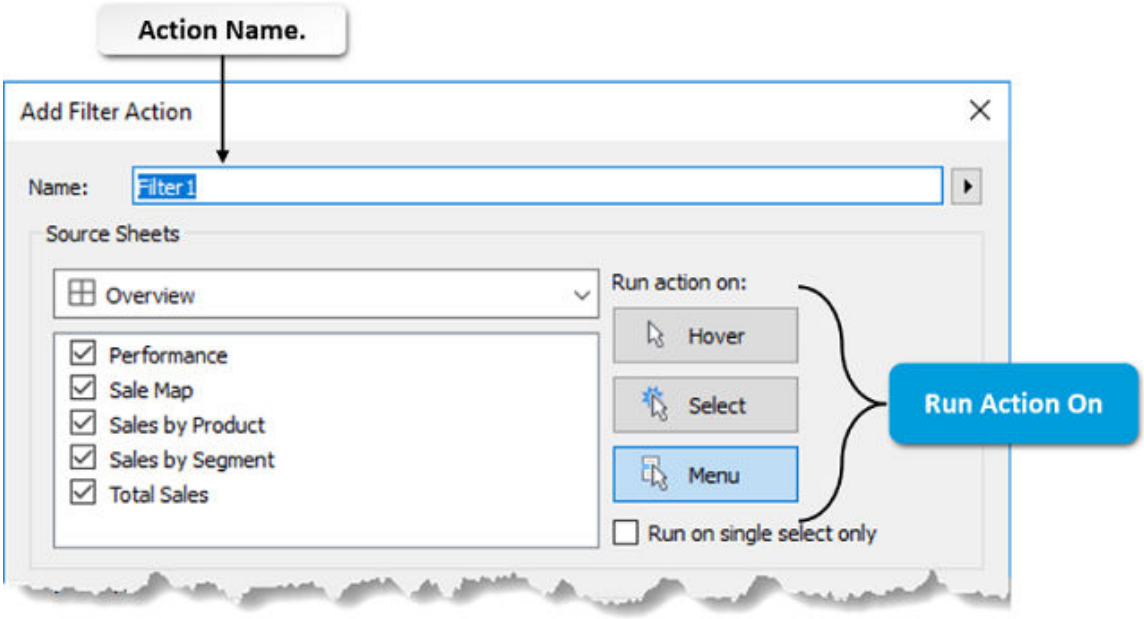



*Figure 8-7: The Name and Run Action On options.*

## Source Sheets and Target Sheets

You must also always configure the source sheet and target sheet. The source sheet is a sheet in the dashboard that will activate the action. All sheets are added by default. You may or may not wish to include all sheets. Typically, you should include sheets with related, relevant data. Marks selected in any of the sheets here will activate the action.

Target sheets specify the dashboard components that the action will apply to. Again, all sheets are added by default. When the action is activated, all of the sheets listed as targets will have the action applied.

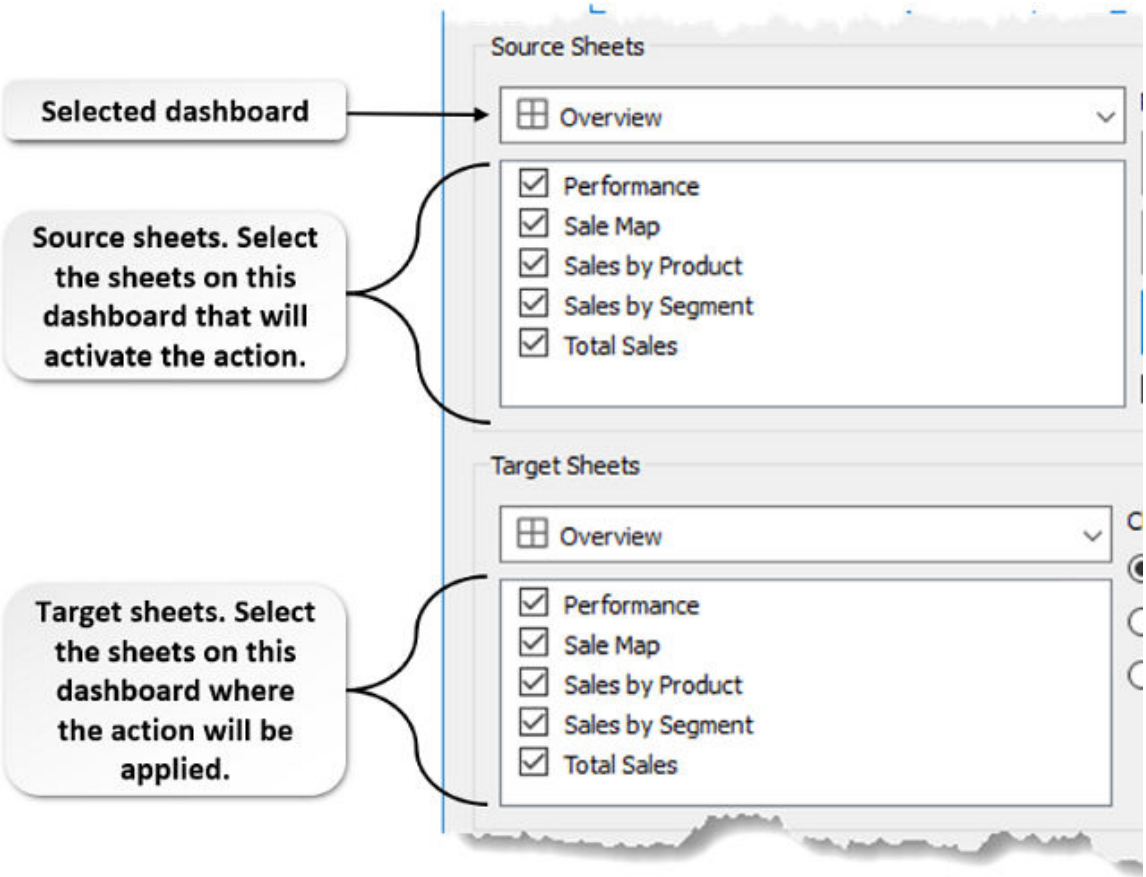



*Figure 8-8: Source Sheets and Target Sheets in a selected dashboard.*

For example, in a dashboard showing sales data for a company by geographic region that includes a map of the regions and sales data in various dashboard components, you might create an action to highlight data from the same region. If all sheets are sources and targets, then selecting the information for a region on any of the dashboard components will highlight data from the same region on other components in the dashboard.

## Filter Action

***Filter actions*** provide more control over filters in the dashboard. For example, if you select one product, the action will filter other dashboard components so only data about the product acted upon is displayed. Filter actions are similar to other filters in Tableau. To create a filter action, from the **Action** menu, select **Add Action→Filter**.

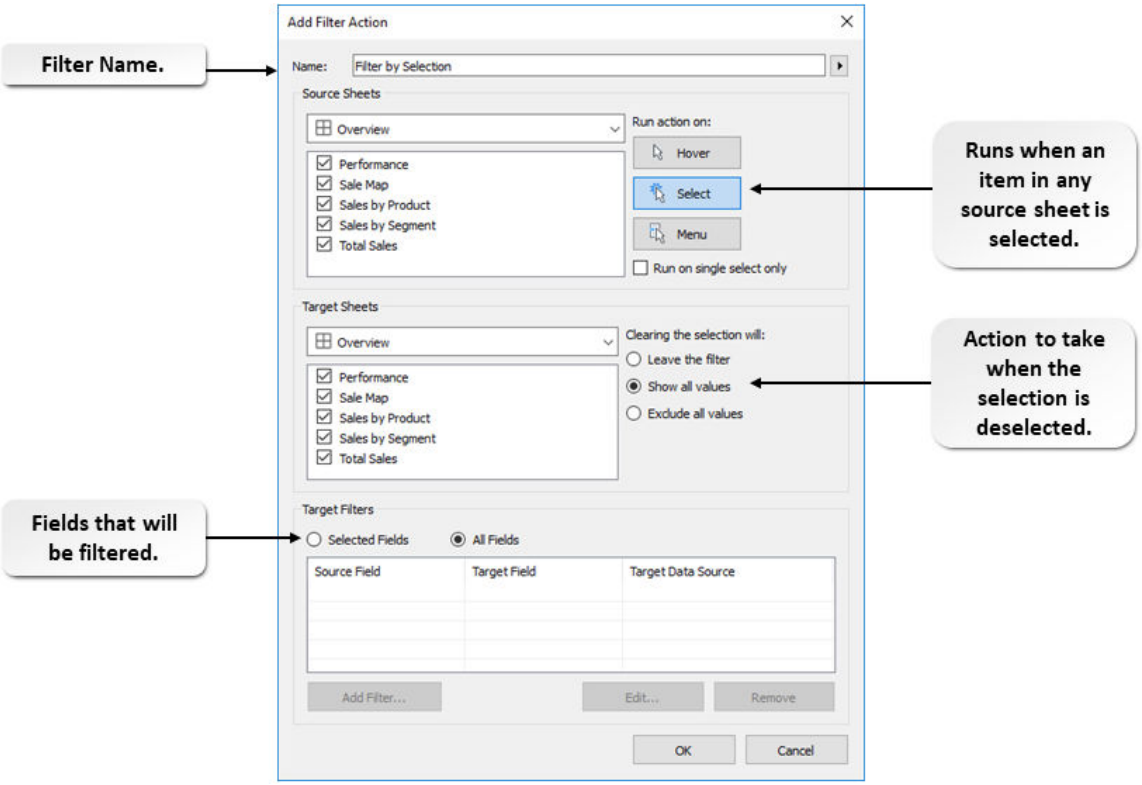



*Figure 8-9: Filter Action options.*

You then must configure the following in addition to the name and run action information.

| *Configuration Object* | *Description* |
|---|---|
| **Clearing the selection will** | Determine what will happen when the selection is cleared in the view. You can opt to **Leave the filter**, **Show all values**, or **Exclude all values**. |
| **Target Filters** | You can select **All Fields** or **Selected Fields**. If you choose **Selected Fields**, you can individually identify source fields, target fields, and target data sources. This is a very powerful feature of Tableau that lets you show different relationships between different fields across different components of the dashboard. |

**Note:** To learn more about using the filter action, check out the Spotlight on **Dashboard Drilldown Reports** presentation from the **Spotlight** tile on the CHOICE Course screen.

## Highlight Action

***Highlight actions*** will highlight all marks in the same category as the mark being acted upon in all dashboard components. These highlight actions are similar to other Tableau highlights. To create a highlight action, from the **Action** menu, select **Add Action→Highlight**.

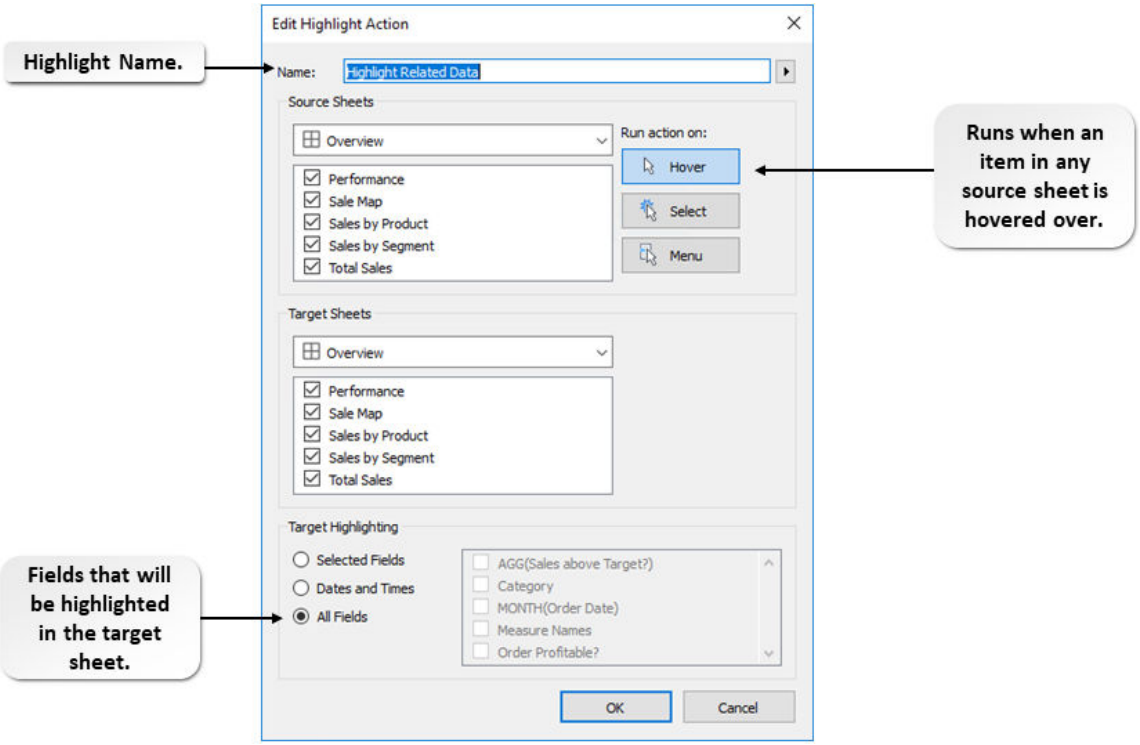



*Figure 8-10: Highlight Action options.*

You then must configure the following in addition to name and run action information.

| *Configuration Object* | *Description* |
|---|---|
| **Target Highlighting** | Select fields from a list, dates, times, or all fields to highlight. |

## URL Action

URL actions let you provide additional information about marks in the view through a link to a web page or file outside of Tableau. As with other actions, field values can be added dynamically to make URLs context sensitive, if the URL supports it.

For example, if you had a map of U.S. cities, you could include a link to the generic URL Wikipedia uses for city pages, and then dynamically add the [City] field to complete the URL. In this case, whenever the link is clicked, it would bring up the Wikipedia page for the currently selected city in the view. You could do the same for internal file-based information. For example, if your dashboard shows product sales information, you could link to internal reports that have additional information for each product, such as the target audience or marketing plan information.

To create a URL action, from the **Action** menu, select **Add Action→URL**.

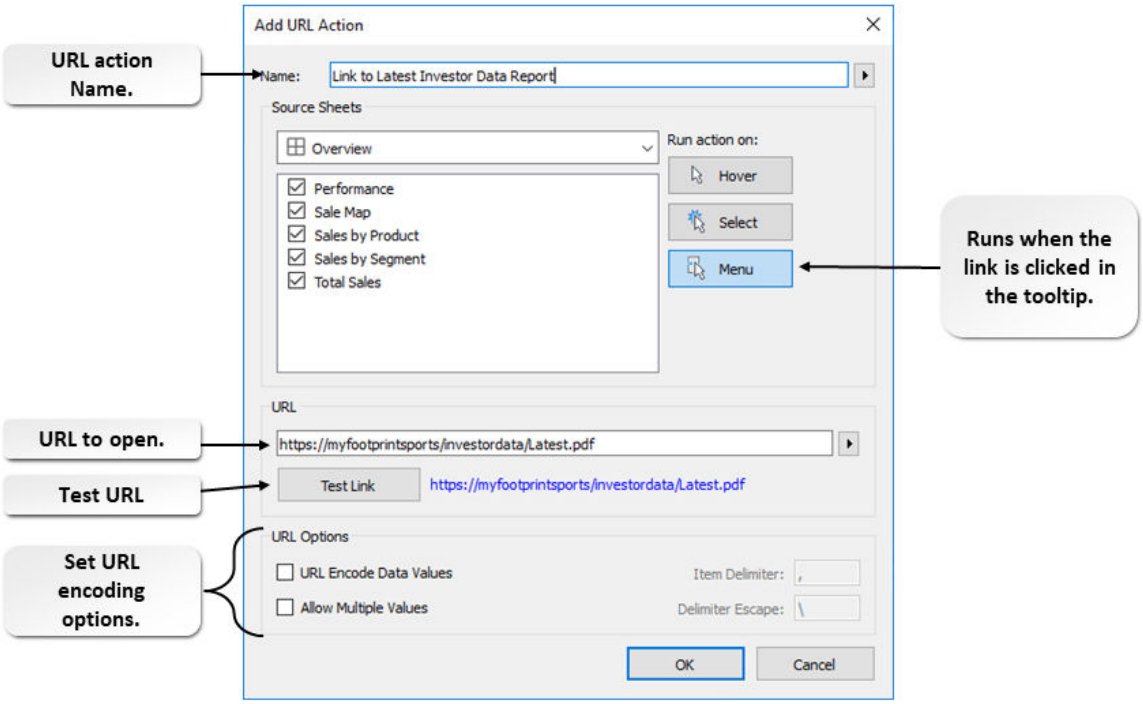



*Figure 8-11: URL Action options.*

You then must configure the following in addition to name, run action, and source sheet information.

| | |
|---|---|
| | **Note:** Since URL actions happen inside the tooltip and don't change the display of other dashboard components, there aren't target sheets to configure. |

| *Configuration Object* | *Description* |
|---|---|
| **URL** | Type the URL or create a URL using dynamic content, such as field names. |
| **Test Link** | Test the URL using data from the view. |
| **URL Options** | You have two options you can configure:<br>• **URL Encode Data Values**: If dynamic data from fields added to the URL contains characters that are not allowed in URLs, check this box to translate those characters through URL encoding.<br>• **Allow Multiple Values**: If the web page you're accessing can accept lists of values, such as a products web page, you can send those parameters by checking this box. You must define a delimiter to separate values and a delimiter escape, which will be used if the delimiter you select is part of the data. |

If you have a web page component in the dashboard, the URL will open in that object in the dashboard. If you do not, the web page will open in your default browser. Web pages open in Tableau are functional and interactive like web pages in a web browser.

## Actions in Tooltips: Run Action on Menu

Adding actions that run on menus provide an excellent way to add interactivity to tooltips. By including URLs or other actions, you add that power and flexibility to the tooltip. It's often helpful

to use actions in tooltips to allow users access to additional sheets with related data or URLs with additional information.

| | |
|---|---|
| | **Access the Checklist tile on your CHOICE Course screen for reference information and job aids on How to Enhance Dashboards with Actions.** |

# ACTIVITY 8-4
## Enhancing Dashboards with Actions

### Before You Begin
The **MyCopy_MyFootPrintSports_2018_8** workbook is open in Tableau Desktop.

### Scenario
You want to enhance the interaction between the views on your sales and profit dashboard. You want to be able to filter the profit view by selecting subcategories in the sales view. You also want to highlight all marks for a subcategory in both views by hovering over a mark for the subcategory in either view. Finally, when comparing shipping method data, you want to be able to review shipping method information on the United States Postal Service website.

---

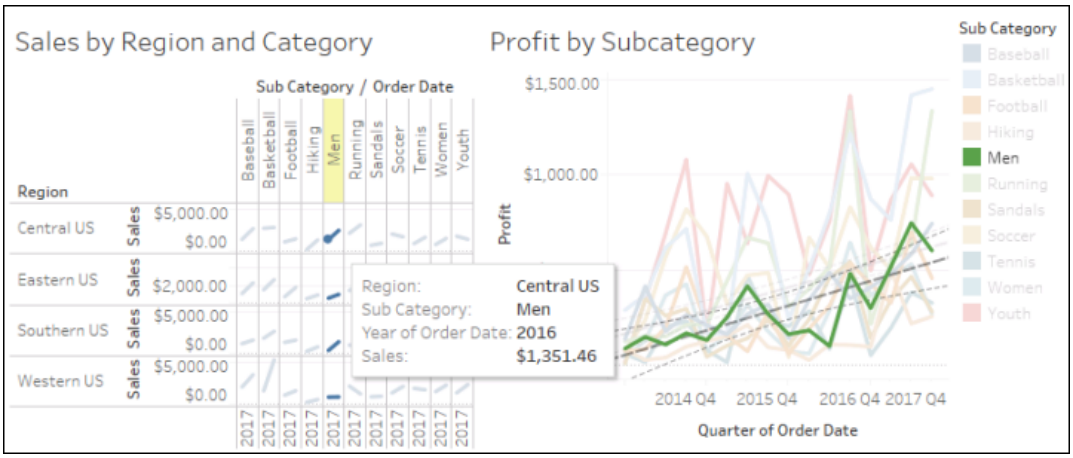
1. Create a filter action.
   a) At the bottom of the screen, select the **Sales/Profit 2016/2017** dashboard.
   b) From the menu, select **Dashboard→Actions**.
   c) In the **Actions** dialog box, select **Add Action→Filter**.
   d) In the **Add Filter Action** dialog box, in the **Name** box, type *Filter Subcategory*
   e) Under **Source Sheets**, uncheck the **Profit by Subcategory** check box.
   f) Under **Run action on**, select **Select**.
   g) Under **Target Sheets**, uncheck the **Sales by Region and Category** check box.
   h) Select **OK**.

2. Create a highlight action.
   a) In the **Actions** dialog box, select **Add Action→Highlight**.
   b) In the **Add Highlight Action** dialog box, in the **Name** box, type **Highlight Subcategory**.
   c) Under **Run action on**, select **Hover**.
   d) Under **Target Highlighting**, select **Selected Fields**.
   e) Check the **Sub Category** check box.
   f) Select **OK**.

3. Create an URL action.
   a) At the bottom of the screen, select the **Shipping 2016/2017** dashboard.
   b) From the menu, select **Dashboard→Actions**.
   c) In the **Actions** dialog box, select **Add Action→Go to URL**.
   d) In the **Add Highlight Action** dialog box, in the **Name** box, type *Shipping Methods*
   e) Under **Run action on**, verify **Menu** is selected.
   f) In the **URL** box, type *https://www.usps.com/ship/mail-shipping-services.htm*
   g) Select **OK** twice.

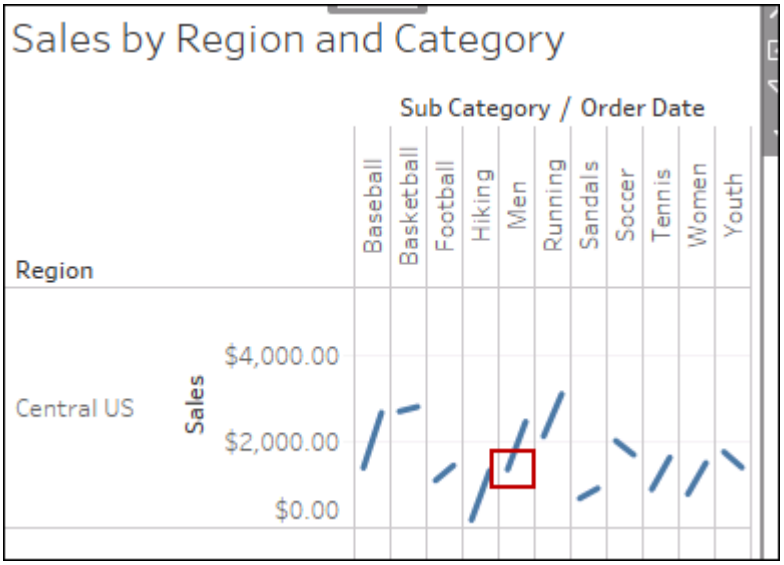
4. Use the highlight action in the story.
   a) At the bottom of the screen, select the **Growth Requires More Staff** story.
   b) At the top of the screen, under **Growth Requires More Staff**, select the first story point.
   c) In the story point, in either view, hover your mouse over different marks.

d) In the dashboard, observe how hovering over a mark in one view will highlight the marks in the other view for the same subcategory.

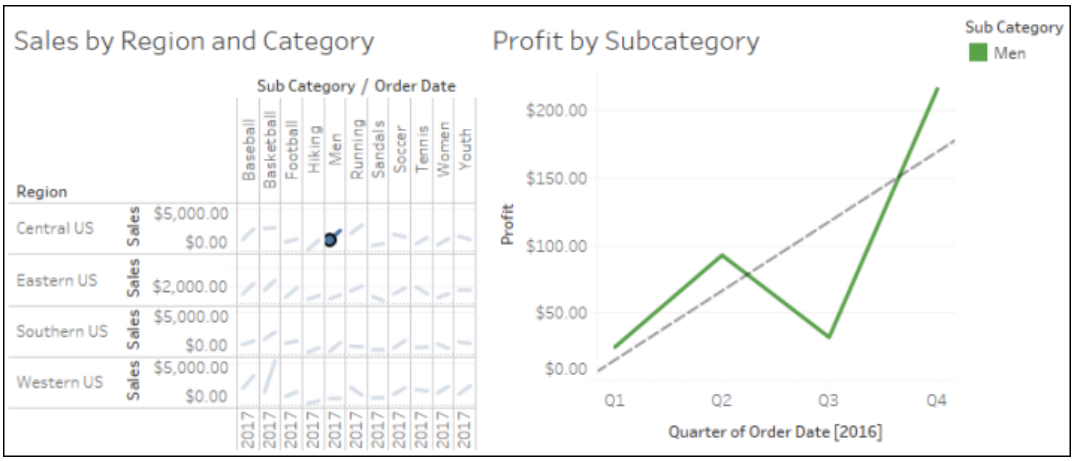



5. Use the filter action in the story.

a) Under **Sales by Region and Category**, in the **Central US** row, for **Men**, select the bottom-left of the line.



b) In the story point, observe the **Sales by Region and Category** view has the **Men** mark highlighted, and that the **Profit by Subcategory** view is filtered to only show the data related to the selected mark.



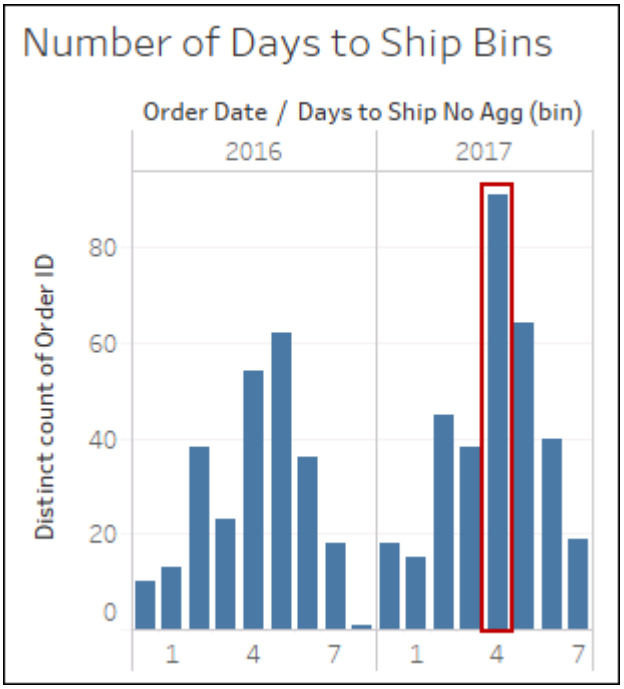

c) Under **Sales by Region and Category**, in the **Central US** row, for **Men**, select the bottom-left of the line to deselect it.

## 6. Use the URL action in the story.

a) At the top of the screen, under **Growth Requires More Staff**, select the second story point.

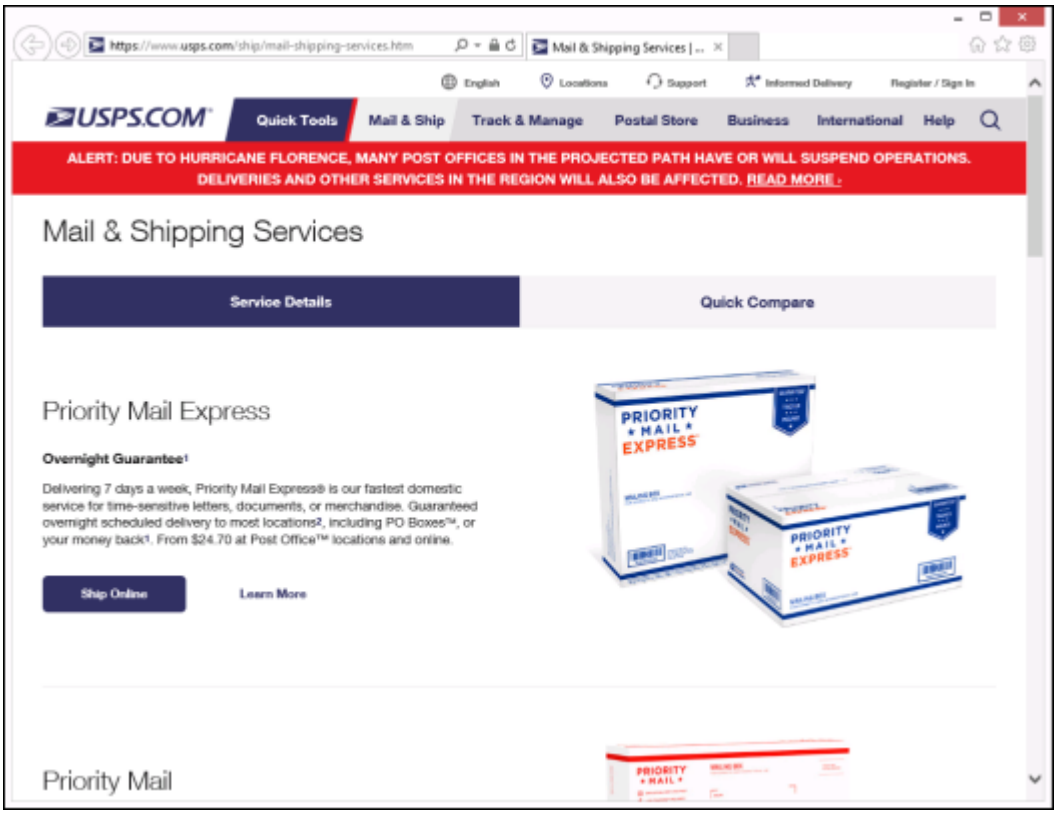
b) In the story point, view **Number of Days to Ship Bins**.

c) For **2017**, select the **4** bar mark and keep the mouse over the mark.




d) In the tooltip, select the **Shipping Methods** link.

e) The USPS Mail & Shipping Services page opens in your default web browser.




f) Close the browser.

7. Save the workbook.

a) In Tableau Desktop, from the menu, select **File→Save**.

b) From the menu, select **File→Close**.

# Summary

In this lesson, you customized mapped data by identifying unrecognized locations, entering custom geocode data, and creating maps with custom background images. You then enhanced visualizations using tooltips to help tell the story of the data and by showing other visualizations inside of the tooltip. Finally, you used dashboard actions to filter and highlight dashboard data, and supply URLs for additional information

## How might you use tooltips to enhance visualizations in your organization?

## What types of actions might you use to enhance dashboard visualizations in your organization?

**Note:** Check your CHOICE Course screen for opportunities to interact with your classmates, peers, and the larger CHOICE online community about the topics covered in this course or other topics you are interested in. From the Course screen you can also access available resources for a more continuous learning experience.